Lifetime Data Anal
DOI 10.1007/s10985-013-9286-0

Local linear estimation of concordance probability with application to covariate effects models on association for bivariate failure-time data

Aidong Adam Ding · Jin-Jian Hsieh · Weijing Wang

Received: 4 May 2012 / Accepted: 26 November 2013

Abstract Bivariate survival analysis has wide applications. In the presence of covariates, most literature focuses on studying their effects on the marginal distributions. However covariates can also affect the association between the two variables. In this article we consider the latter issue by proposing a nonstandard local linear estimator for the concordance probability as a function of covariates. Under the Clayton copula, the conditional concordance probability has a simple one-to-one correspondence with the copula parameter for different data structures including those subject to independent or dependent censoring and dependent truncation. The proposed method can be used to study how covariates affect the Clayton association parameter without specifying marginal regression models. Asymptotic properties of the proposed estimators are derived and their finite-sample performances are examined via simulations. Finally, for illustration, we apply the proposed method to analyze a bone marrow transplant data set.

Keywords Multivariate local polynomial regression · Clayton copula · Non-informative missing data · Dependent censoring · Dependent truncation

1 Introduction

Copula models have been a popular choice for modeling multivariate failure-time data due to the nice feature that the dependence structure can be studied separately from the

A. A. Ding (✉)
Department of Mathematics, Northeastern University, Boston, MA 02115, USA
e-mail: a.ding@neu.edu

J.-J. Hsieh
Department of Mathematics, National Chung Cheng University, Chia-Yi, Taiwan, ROC

W. Wang
Institute of Statistics, National Chiao-Tung University, Hsin-Chu, Taiwan, ROC

Published online: 10 December 2013

marginal distributions. Consider the bivariate case. Let T_1 and T_2 be a pair of lifetime variables with continuous marginal distribution (or survival) functions denoted as F_1 and F_2, respectively. Sklar's theorem ensures that there exists a unique copula function $C : [0, 1]^2 \to [0, 1]$ such that the corresponding joint function F can be written as $F(t_1, t_2) = C\{F_1(t_1), F_2(t_2)\}$. The dependence structure is characterized by the copula function $C(\cdot, \cdot)$.

Semi-parametric inference for parametric versions of C has been extensively studied for right censored data without covariates (Clayton 1978; Oakes 1989; Shih and Louis 1995; Wang and Wells 2000). In recent years, applications of copula models have been extended to more complicated data structures, including semi-competing risks data (Day et al. 1997; Fine et al. 2001; Wang 2003; Jiang et al. 2005; and Lakhal et al. 2008) and dependent truncation data (Chaieb et al. 2006; Emura and Wang 2010) under which the copula structure also helps to resolve the problem of non-identifiability in marginal inference.

When covariates Z are present, they may affect the dependence structure. The dependence structure conditional on $Z = z$ can be described by the conditional copula C_z. Gijbels et al. (2011) proposed a nonparametric estimator for the conditional copula based on kernel-smoothing of the bivariate and marginal empirical distributions. Acar et al. (2011) proposed another kernel-based method for estimating the conditional copula association parameter using the local likelihood approach (Fan and Gijbels 1996). Both estimations are based on complete data from a bivariate joint distribution. Their methods are two-stage procedures: first estimate the marginals, then estimate the copula parameters using the marginal estimates from the first stage. However those approaches cannot be extended to semi-competing risks data and dependent truncation data in which the marginal functions are not identifiable nonparametrically. In this paper, we propose a general approach to estimating the conditional concordance probability for various data structures. Under the Clayton copula model, this method provides an estimator for the association parameter as a function of the covariates applicable to various data structures.

When one failure time is subject to dependent censoring by the other time variable, the data structure is called semi-competing risk data. It turns out that one marginal distribution is not identifiable nonparametrically. Ding (2010) studied the identifiability condition when the copula assumption is imposed on the joint distribution. Most existing methods for analyzing this data structure make the implicit assumption that covariates have no effect on the dependence structure. The focus is mostly on marginal regression analysis (Lin et al. 1996; Lin and Ying 2003; Peng and Fine 2006; Huang and Zhang 2008; Ding et al. 2009). Some papers (Fine and Jiang 2000; Peng and Fine 2007) proposed to estimate the copula parameter assuming that it is constant across different covariate values. Ghosh (2006) proposed to test constancy of association across discrete covariate strata. Hsieh and Wang (2008) considered marginal regression analysis in which the dependence structure can vary only for discrete covariate groups. Similar issues also arise in dependent truncation data (Ding 2012).

The strength of association for a copula C_z can be measured by Kendall's tau $\tau(z) = 4\int\int C_z(u, v)dC_z(u, v) - 1$, or equivalently the concordance probability $\alpha(z) = (\tau(z) + 1)/2$ between pairs of the bivariate failure times. We propose to estimate $\alpha(z)$ by a local linear estimator using concordance indicators as the response variables in regression. While properties of multivariate local polynomial regression have been well studied in literature (e.g., Ruppert and Wand 1994; Opsomer and Ruppert 1997), the problem is fundamentally different from the usual local polynomial regression setting. Here the response variables are the concordance indicators for pairs of observations and hence are no longer mutually independent. Also, for the data structures which involve censoring or truncation, not all the bivariate failure times are observed. Accordingly the concordance relationship between some pairs is not observable either. In these cases, we modify the proposed local linear estimator to estimate the conditional concordance probability for comparable pairs denoted as $\alpha^*(z)$. For dependent truncation data, $2\alpha^* - 1$ is the conditional Kendall's tau studied in Tsai (1990), Martin and Betensky (2005) and Emura and Wang (2010).

The missing-at-random mechanism is the key property to further establish $\alpha^*(z) = \alpha(z)$ for the data structures in the presence of censoring and truncation. The Clayton copula has a special feature that the strength of association remains constant across time. We will show that the Clayton assumption leads to a missing-at-random property which makes $\alpha^*(z) = \alpha(z)$ for all the data structures in the presence of dependent censoring and truncation considered here. Without covariate effects, several papers including Fine and Jiang (2000), Fine et al. (2001) and Jiang et al. (2005) utilized the concordance information between pairs of the two variables to estimate the Clayton association parameter. Our estimator provides a unified approach to estimating the Clayton association parameter applicable to various data structures including dependent censoring and dependent truncation without specifying the marginal regression models.

The paper is organized as follow. In Sect. 2, we define the model assumption and four common data structures including those with censoring and truncation. The asymptotic properties of the proposed estimator are discussed in Sect. 3. We then evaluate the estimator through simulation studies and apply it to a bone marrow transplant data set for illustration.

2 Model assumption and data structures

2.1 Local linear estimation for concordance probability

The copula representation of the bivariate joint function of (T_1, T_2) conditional on covariate value $Z = z$ can be written as

$$F_z(t_1, t_2) = C_{\theta(z)}\{F_{1,z}(t_1), F_{2,z}(t_2)\},$$

where $F_{1,z}(\cdot)$ and $F_{2,z}(\cdot)$ are the conditional marginal functions. We assume that Z is a d-dimensional vector of covariates.

Let $(T_{1,i}, T_{2,i})$ be independently and identically distributed $(i.i.d.)$ random replications of (T_1, T_2). Define $\delta_{ij} = I\{(T_{1,i} - T_{1,j})(T_{2,i} - T_{2,j}) > 0\}$ as the concordance indicator for $(T_{1,i}, T_{2,i})$ and $(T_{1,j}, T_{2,j})$ $(i \neq j)$. Then the concordance probability conditional on $Z_i = Z_j = z$ is $\alpha(z) = \Pr(\delta_{ij} = 1|Z_i = Z_j = z)$. Note that for a copula model, $\alpha(z) = 2\int\int C_{\theta(z)}(u, v)dC_{\theta(z)}(u, v)$. We propose to estimate $\alpha(z)$ by the local (in z) linear estimator that minimizes

$$\sum_{i<j}[\delta_{ij} - \alpha(z) - \beta_1^T(Z_i - z) - \beta_2^T(Z_j - z)]^2 K_H[(Z_i - z, Z_j - z)^T]$$

over $\alpha(z)$ and $\beta = (\beta_1^T, \beta_2^T)^T$, where

$$\begin{aligned} K_H[(u_1, u_2)^T] &= [|H|^{-1}K(H^{-1}u_1)][|H|^{-1}K(H^{-1}u_2)] \\ &= |H|^{-2}K(H^{-1}u_1)K(H^{-1}u_2), \end{aligned}$$

and the determinant $|H| \to 0$ when $n \to \infty$. Denote the resulting estimator as $\hat{\alpha}(z)$. Notice that the $2d$-dimensional kernel $K_H(\cdot)$ is a d-dimensional kernel K multiplied by itself and the $2d$-dimensional bandwidth matrix is

$$\begin{pmatrix} H & 0 \\ 0 & H \end{pmatrix}$$

with H being d-dimensional. Notice that our setting is different from the standard $2d$-dimensional multivariate nonparametric regression in that Z_i and Z_j are actually measurements of the same variable from different subjects. Hence we use the same bandwidth matrix H for Z_i and Z_j due to symmetry.

In the presence of censoring or truncation, δ_{ij} may not be observable for some (i, j) pairs, which can be viewed as a missing data phenomenon. To tackle the problem, we first consider estimating the concordance probability for a subset of the sample with δ_{ij} observable for all (i, j) pairs in the subset. Specifically, define Δ_{ij} as the indicator that the pair (i, j) belongs to the target subset. Let $\tilde{\alpha}(z_1, z_2) = E[\delta_{ij}|\Delta_{ij} = 1, Z_i = z_1, Z_j = z_2]$. Then the local linear estimator estimates the conditional probability $E(\delta_{ij}|\Delta_{ij} = 1, Z_i = Z_j = z) = \tilde{\alpha}(z, z)$ by minimizing

$$\sum_{i<j}[\delta_{ij} - \alpha(z) - \beta_1^T(Z_i - z) - \beta_2^T(Z_j - z)]^2 \Delta_{ij} K_H[(Z_i - z, Z_j - z)^T] \tag{1}$$

over $\alpha(z)$ and $\beta = (\beta_1^T, \beta_2^T)^T$.

A crucial missing-at-random condition is

$$E(\delta_{ij}|\Delta_{ij} = 1, Z_i = Z_j = z) = \Pr(\delta_{ij} = 1|Z_i = Z_j = z) = \alpha(z). \tag{2}$$

When condition (2) holds, $\hat{\alpha}(z)$ estimates the quantity $\tilde{\alpha}(z, z) = \alpha(z)$ which has the natural correspondence with the copula parameter $\theta(z)$. Without condition (2), we may still use $\hat{\alpha}(z)$ to examine covariate effect which however cannot be easily converted to the copula parameter $\theta(z)$. We will show that condition (2) holds for some common data structures subject to censoring and truncation when the Clayton copula assumption is made. Thus the estimator $\hat{\alpha}(z)$ provides a unified approach to estimating the covariate effect on the Clayton association parameter in these data structures.

2.2 Four data structures

Now we discuss four different data structures that (1) can be applied with appropriate construction of Δ_{ij}.

Data structure 1: bivariate failure-time data without censoring. We observe an $i.i.d.$ sample $(T_{1,i}, T_{2,i}, Z_i)$ $(i = 1, ..., n)$ from (T_1, T_2). Thus $\Delta_{ij} = 1$ for all i, j.

Data structure 2: bivariate failure-time data with independent right censoring. Denote $(Y_{1,i}, Y_{2,i})$ $(i = 1, \ldots, n)$ as a random sample from the bivariate censoring variables, (Y_1, Y_2). It is assumed that (T_1, T_2) are independent of (Y_1, Y_2) conditional on $Z = z$. Under right censoring, one observes $(X_{1,i}, X_{2,i}, \tilde{\delta}_{1,i}, \tilde{\delta}_{2,i}, Z_i)$ $(i = 1, \ldots, n)$, where $X_{k,i} = T_{k,i} \wedge Y_{k,i}$ and $\tilde{\delta}_{k,i} = I(T_{k,i} \leq Y_{k,i})$ for $k = 1, 2$ and $x \wedge y$ denotes the minimum of x and y. In this case, $\Delta_{ij} = 1$ if and only if $T_{k,i} \wedge T_{k,j} < Y_{k,i} \wedge Y_{k,j}$ for both $k = 1, 2$. That is, $X_{k,i} \wedge X_{k,j}$ corresponds to an uncensored point for each $k = 1, 2$.

Data structure 3: semi-competing risks data with independent right censoring. Consider the situation that T_{1i} is subject to dependent censoring by T_{2i} and both are subject to independent censoring by Y_i, an independent replication from Y. It is assumed that (T_1, T_2) are independent of Y conditional on $Z = z$. Observed variables can be written as $(X_{1,i}, X_{2,i}, \tilde{\delta}_{1,i}, \tilde{\delta}_{2,i}, Z_i)$ $(i = 1, \ldots, n)$, where $X_{1,i} = T_{1,i} \wedge T_{2,i} \wedge Y_i$, $X_{2,i} = T_{2,i} \wedge Y_i$, $\tilde{\delta}_{1,i} = I\{X_{1,i} = T_{1,i}\}$ and $\tilde{\delta}_{2,i} = I\{X_{2,i} = T_{2,i}\}$. In this case, $\Delta_{ij} = 1$ if and only if $T_{1,i} \wedge T_{1,j} < T_{2,i} \wedge T_{2,j} < Y_i \wedge Y_j$. That is, $X_{k,i} \wedge X_{k,j}$ corresponds to an uncensored point for each $k = 1, 2$.

Data structure 4: dependent left truncation data with independent right censoring. The variable T_{2i} is subject to left truncation by T_{1i} and independent right censoring by Y_i, an independent replication from Y. It is assumed that (T_1, T_2) are independent of Y conditional on $Z = z$. One observes $(T_{1,i}, X_{2,i}, \tilde{\delta}_{2,i}, Z_i)$ $(i = 1, \ldots, n)$ subject to $T_{1,i} \leq X_{2,i}$, where $X_{2,i} = T_{2,i} \wedge Y_i$, $\tilde{\delta}_{2,i} = I\{X_{2,i} = T_{2,i}\}$. In this case, $\Delta_{ij} = 1$ if and only if $T_{1,i} \vee T_{1,j} < X_{2,i} \wedge X_{2,j}$ and $T_{2,i} \wedge T_{2,j} < Y_i \wedge Y_j$, where $x \vee y$ denotes the maximum of x and y.

2.3 Clayton copula and condition (2)

For the first data structure, without censoring or truncation, $\Delta_{ij} = 1$ for sure and obviously the condition (2) holds for all copulas. For the second and third data structures with censoring, we assume the Clayton copula of the form:

$$C_{\theta(z)}(u_1, u_2) = (u_1^{-\theta(z)} + u_2^{-\theta(z)} - 1)^{-1/\theta(z)} \quad (\theta(z) > 0).$$

For the fourth data structure, we assume the extended-Clayton copula of the form:

$$\tilde{C}_{\theta(z)}(u_1, u_2) = u_1 - (u_1^{-\theta(z)} + (1 - u_2)^{-\theta(z)} - 1)^{-1/\theta(z)} \quad (\theta(z) > 0).$$

Imposing the extended-Clayton copula on the function $F(t_1, t_2) = Pr(T_1 \leq t_1, T_2 \leq t_2)$ is equivalent to imposing a Clayton copula on $\tilde{F}(t_1, t_2) = Pr(T_1 \leq t_1, T_2 > t_2)$.

For dependent truncation data, mathematically it is better to impose a copula on $\tilde{F}(t_1, t_2)$ (Chaieb et al. 2006; Emura and Wang 2010).

In Appendix 1, we verify Eq. (2) based on these data structures. For continuous survival times T_1 and T_2, under the Clayton copula, $\theta(z) = [2\alpha(z) - 1]/[1 - \alpha(z)]$. Therefore, estimation of $\theta(z)$ is equivalent to that of $\alpha(z)$.

3 Asymptotic properties

To derive the asymptotic properties and variance formula of the proposed estimator, we adopt similar notations as in Ruppert and Wand (1994) for multivariate local polynomial regression. Specifically denote $\boldsymbol{\delta} = (\delta_{ij})_{i<j}$, a $n(n-1)/2$-dimensional vector, as the response. The input variable for δ_{ij} is $Z_{ij} = (Z_i^T, Z_j^T)^T$ where v^T denotes the transpose of vector v. Since the conditional expectation of δ_{ij} is $\tilde{\alpha}(Z_{ij})$, we are working on a regression problem with δ_{ij} being the response variable and the $2d$-dimensional vector Z_{ij} being the explanatory variables. However, the problem is different from the standard setting for $(2d)$-dimensional multivariate local polynomial regression. First, we are only interested in estimating the d-dimensional function $\tilde{\alpha}(z, z)$ instead of the $2d$-dimensional function $\tilde{\alpha}(z_1, z_2)$. Secondly, the response δ_{ij} and input Z_{ij} are not independent replicates as in the usual regression setting. Specifically (δ_{ij}, Z_{ij}) $(i < j)$ are not independent among different (i, j) pairs. Finally the value of some δ_{ij} may be unknown. As discussed earlier, we will select pairs with $\Delta_{ij} = 1$ in the analysis so that the corresponding δ_{ij} is always observable.

To simplify the notations, we use $\mathbf{z} = (z^T, z^T)^T$. For example, $Z_{ij} - \mathbf{z}$ denotes $((Z_i - z)^T, (Z_j - z)^T)^T$, and $\alpha(z) = \tilde{\alpha}(z, z) = \tilde{\alpha}(\mathbf{z})$. So Eq. (1) can also be written as $\sum_{i<j}[\delta_{ij} - \tilde{\alpha}(\mathbf{z}) - \beta^T(Z_{ij} - \mathbf{z})]^2 \Delta_{ij} K_H(Z_{ij} - \mathbf{z})$. Denote the gradient of $\tilde{\alpha}(\mathbf{z})$ as

$$\mathcal{D}_\alpha(z) = \begin{pmatrix} \frac{\partial}{\partial z_1}\tilde{\alpha}(z_1, z_2) \\ \frac{\partial}{\partial z_2}\tilde{\alpha}(z_1, z_2) \end{pmatrix}_{z_1 = z_2 = z},$$

and the Hessian matrix of $\tilde{\alpha}(\mathbf{z})$ as

$$\mathcal{H}_\alpha(z) = \begin{pmatrix} \frac{\partial^2}{\partial z_1 \partial z_1^T}\tilde{\alpha}(z_1, z_2) & \frac{\partial^2}{\partial z_1 \partial z_2^T}\tilde{\alpha}(z_1, z_2) \\ \frac{\partial^2}{\partial z_2 \partial z_1^T}\tilde{\alpha}(z_1, z_2) & \frac{\partial^2}{\partial z_2 \partial z_2^T}\tilde{\alpha}(z_1, z_2) \end{pmatrix}_{z_1 = z_2 = z}.$$

Properties about the asymptotic bias and variance of the estimator $\hat{\alpha}(z)$ are summarized in the following theorem. The proof is provided in Appendix 2.

Theorem 1 *We assume the following technical conditions similar to those in Ruppert and Wand (1994).*

(i) The kernel function K is a compactly supported, bounded non-negative kernel such that $\int K(u)du = 1$, $\int K(u)uu^T du = \mu_2(K) I_d$ where $\mu_2(K) > 0$ is a scalar and I_d is the $d \times d$ identity matrix. And all odd-order moments of K vanish.

(ii) When $n \to \infty, n|H| \to \infty$, and all entries in the symmetric positive definite matrix H tend to zero. Also, the condition number of H is bounded.

(iii) Denote f_Z as the density of Z. Assume that z is an interior point of the support of f_Z with $f_Z(z) > 0$.

(iv) Denote the non-missing probability function $\tilde{g}(z_1, z_2) = E[\Delta_{12}|Z_1 = z_1, Z_2 = z_2]$ and $\breve{g}(z_1, z_2, z_3) = E[\Delta_{12}\Delta_{13}|Z_1 = z_1, Z_2 = z_2, Z_3 = z_3]$. Assume $\tilde{g}(z, z) > 0, \breve{g}(z, z, z) > 0$ and $0 < \alpha(z) < 1$.

Then

$$E[\hat{\alpha}(z)|Z_1, ..., Z_n] = \alpha(z) + \mu_2(K)tr\{H^2\mathcal{H}_\alpha(z)\} + o_p(tr(H^2)) \tag{3}$$

and

$$Var[\hat{\alpha}(z)|Z_1, ..., Z_n] = \frac{4(n-2)\breve{g}(z, z, z)\breve{\alpha}(z, z, z)\mu_0(K^2)}{n(n-1)|H|[\tilde{g}(z, z)]^2 f_Z(z)} + o_p(n^{-1}|H|^{-1}), \tag{4}$$

where $tr(M)$ denotes the trace of matrix M, $\mu_0(K^2) = \int[K(u)]^2du$ and

$$\begin{aligned}\breve{\alpha}(z_1, z_2, z_3) &= Cov[\delta_{12}, \delta_{13}|\Delta_{12} = \Delta_{13} = 1, Z_1 = z_1, Z_2 = z_2, Z_3 = z_3]\\ &= E[\delta_{12}\delta_{13}|\Delta_{12} = \Delta_{13} = 1, Z_1 = z_1, Z_2 = z_2, Z_3 = z_3]\\ &\quad -\tilde{\alpha}(z_1, z_2)\tilde{\alpha}(z_1, z_3).\end{aligned}$$

To better understand the above results, let us consider the simple case of a univariate Z. Then the scalar $H = h \to 0$. The above theorem states that $Bias(\hat{\alpha}|Z_1, ..., Z_n) = O_p(h^2)$ and $Var(\hat{\alpha}|Z_1, ..., Z_n) = O_p(n^{-1}h^{-1})$. So the estimator has the best rate of convergence when $O_p(h^4) = O_p(n^{-1}h^{-1})$. That is $h = n^{-1/5}$ and the rate of convergence for $\hat{\alpha}(z)$ is $n^{-2/5}$.

Remark 1 Theorem 1 states the property of $\hat{\alpha}(z)$ which estimates $\alpha^*(z) = E(\delta_{ij}|\Delta_{ij} = 1, Z_i = Z_j = z)$. Under condition (2), $\alpha^*(z)$ agrees with $\alpha(z) = \Pr(\delta_{ij} = 1|Z_i = Z_j = z)$. For complete data, since condition (2) always holds, $\hat{\alpha}(z)$ can be directly translated to estimate $\theta(z)$ for any specified one-parameter copula family such as Frank or Gumbel models (Nelsen 2006). For other three data structures, we need to assume the Clayton copula under which the missing-at-random condition in (2) holds. In these complicated data structures, $\hat{\alpha}(z)$ can still estimate Clayton's association parameter.

Remark 2 Theorem 1 shows that the fitted estimate $\hat{\alpha}(z)$ converges to the true value $\alpha(z)$ for $0 < \alpha(z) < 1$. It may happen that some fitted values of $\hat{\alpha}(z)$ fall outside the unit interval [0, 1] especially when the true $\alpha(z)$ is near 0 or 1. Such a phenomenon occurs rarely in our numerical studies. When they do happen, we can truncate these estimates to make them equal to the nearest boundary value 0 or 1. Another possible approach to dealing with this issue is to incorporate a link function $g(\cdot) : (-\infty, \infty) \to (0, 1)$ such that $\gamma(z) = g^{-1}(\alpha(z))$ is unbounded. Then we estimate $\gamma(z)$ by minimizing $\sum_{i<j}[\delta_{ij} - g[\gamma(z)] - g'[\gamma(z)]\beta_1^T(Z_i - z) - g'[\gamma(z)]\beta_2^T(Z_j - z)]^2 K_H(Z_{ij} - z)$. This alternative extension is a topic for future exploration.

The boundary effect is often a concern for smoothing techniques. Now we investigate whether the asymptotic bias and variance remain the same order at the boundary of the support $supp(f_Z)$ for covariate Z. As in Ruppert and Wand (1994), we consider a sequence $z = z_n$ converging to a point z_∂ on the boundary of $supp(f_Z)$. That is, $z = z_\partial + H^{1/2}c$ for a fixed c. We also assume the following condition to avoid degeneracy.

(v) There is a convex set $\mathfrak{C}$ with nonnull interior and containing z_∂ such that

$$\inf_{z\in\mathfrak{C}} f_Z(z) > 0. \tag{5}$$

At the boundary point $z = z_\partial + H^{1/2}c$, the effective support of kernel is reduced. Let $\mathcal{D}_{z,H} = \{x : (z + H^{1/2}x) \in supp(f_Z)\} \cap supp(K)$. We denote $\mu^*_{z,0}(K) = \int_{\mathcal{D}_{z,H}} K(u)du$, $\mu^*_{z,1}(K) = \int_{\mathcal{D}_{z,H}} K(u)udu$, $\mu^H_{z,2}(K) = \int_{\mathcal{D}_{z,H}} K(u)u^T H\mathcal{H}_\alpha(z)Hudu = O(H^2)$ and $\mu^H_{z,3}(K) = \int_{\mathcal{D}_{z,H}} uK(u)u^T H\mathcal{H}_\alpha(z)Hudu = O(H^2)$. Also let

$$N_z = \int\limits_{\mathcal{D}_{z,H}} \int\limits_{\mathcal{D}_{z,H}} (1\ u_1^T\ u_2^T)^T(1\ u_1^T\ u_2^T)K(u_1)K(u_2)du_1du_2,$$

$$T_z = \int\limits_{\mathcal{D}_{z,H}} \int\limits_{\mathcal{D}_{z,H}} \int\limits_{\mathcal{D}_{z,H}} (1\ u_1^T\ u_2^T)^T(1\ u_1^T\ u_2^T)K(u_1)K(u_2)K^2(u_3)du_1du_2du_3.$$

Theorem 2 *Suppose that $z = z_\partial + H^{1/2}c$ for a fixed $c \in supp(K)$. We assume conditions (i)- (iv) as in Theorem 1 and condition (v) in Eq. (5). Then*

$$Bias[\hat{\alpha}(z)|Z_1, ..., Z_n] = \frac{e_1^T N_z^{-1}}{2} \begin{pmatrix} 2\mu^*_{z,0}(K)\mu^H_{z,2}(K) \\ \mu^H_{z,2}(K)\mu^*_{z,1}(K) + \mu^*_{z,0}(K)\mu^H_{z,3}(K) \\ \mu^H_{z,2}(K)\mu^*_{z,1}(K) + \mu^*_{z,0}(K)\mu^H_{z,3}(K) \end{pmatrix} + o_p(tr(H^2)) \tag{6}$$

and

$$Var[\hat{\alpha}(z)|Z_1, ..., Z_n] = \frac{4\check{g}(z,z,z)\check{\alpha}(z,z,z)e_1^T N_z^{-1}T_zN_z^{-1}e_1}{n|H|[\tilde{g}(z,z)]^2 f_Z(z)}[1 + o_p(1)]. \tag{7}$$

The proof of Theorem 2 is provided in Appendix 6. Theorem 1 and 2 show that the conditional bias is of the same order $O_p(tr(H^2))$ at the interior as well as the boundary. The conditional variance also remain the same order $O_p(n^{-1}|H|^{-1})$ at the boundary. Therefore asymptotically the proposed estimator does not suffer from the boundary effect. This result is similar to that of Ruppert and Wand (1994). However the finite sample performances, in particular the variance, can still be affected near the boundary since this region contains fewer observations.

4 Simulation studies

We examine the finite-sample performances of the proposed estimator and compare it with the estimator proposed by Acar et al. (2011). The data generation scheme is similar to that in Acar et al. (2011) for the purpose of comparison. Specifically the covariate Z_i is generated independently from $Uniform(2, 5)$ for $i = 1, \dots, n$ and, given Z_i, $(U_{1,i}, U_{2,i})$ are generated from the Clayton copula $C_{\theta(Z_i)}(u_1, u_2)$. Two forms of $\theta(z)$ are considered: (1) linear calibration function $\theta(z) = \exp(0.8z - 2)$ and (2) Quadratic calibration function: $\theta(z) = \exp(2 - 0.3(z - 4)^2)$. Then we set $T_{1,i} = -0.5 \exp(\gamma_1 Z_i) \log(U_{1,i})$ and $T_{2,i} = -\exp(\gamma_2 Z_i) \log(U_{2,i})$. This means that the marginal distributions are exponentially distributed and $T_{k,i}$ follows the accelerated failure times (AFT) model with parameter γ_k such that $\log(T_{k,i}) = \gamma_k Z_i + e_{k,i}$ for $k = 1, 2$. We set $n = 100$ and replications=400. Throughout the section, we use the Epanechnikov kernel function $K(x) = \max\{0, 0.75(1 - x^2)\}$.

We first assess the situation in the absence of censoring and truncation and compare our estimator with the Acar-Craiu-Yao estimator which is obtained by maximizing a local likelihood function based on a random sample of (U_1, U_2). Here since only $(T_{1,i}, T_{2,i}, Z_i)$ are observed, $(U_{1,i}, U_{2,i})$ need to be estimated. We fit the AFT model on $T_{1,i}$'s and $T_{2,i}$'s to obtain estimates of $(\hat{U}_{1,i}, \hat{U}_{2,i})$ and then carry out the Acar-Craiu-Yao procedure. We also study the effect of mis-specifying the marginal models on the Acar-Craiu-Yao estimator by fitting marginal location shift (LOC) models, $T_{k,i} = \gamma_k Z_i + e_{k,i}$ for $k = 1, 2$. Acar et al. (2011) suggested a cross-validation rule to select the bandwidth h from 12 candidate values, ranging from 0.33 to 2.96, equally spaced on the logarithm scale. Since the proposed procedure is developed to work for all four data structures, the bandwidth selection criterion should not rely on the local likelihood. Accordingly we choose the value of h that minimizes the cross-validation criterion

$$\sum_{i=1}^{n} \sum_{j \neq i} \Delta_{ij} [\delta_{ij} - \hat{\alpha}_{-(i,j)}(z_i, z_j)]^2, \tag{8}$$

where $\hat{\alpha}_{-(k,l)}(z_k, z_l)$ is the estimate from minimizing

$$\sum_{i<j} [\delta_{ij} - \tilde{\alpha}(z_k, z_l) - \beta_1^T (Z_i - z_k) - \beta_2^T (Z_j - z_l)]^2 \Delta_{ij} K_H[(Z_i - z, Z_j - z)^T]$$

without using the concordance of $(k, l)th$ pair.

As in Acar et al. (2011), the performances based on Kendall's $\tau(z) = 2\alpha(z) - 1$ are reported. Three accuracy measures are evaluated: the average squared bias (ABIAS^2), the average variance (AVAR) and the average mean square error (AMSE) denoted as

$$ABIAS^2(\hat{\tau}) = \frac{1}{b - a} \int_a^b \{E[\hat{\tau}(z)] - \tau(z)\}^2 dz,$$

Table 1 Estimation of $\tau(z)$ without censoring and truncation under marginal AFT models (data structure 1)

	(γ_1, γ_2) = (0,0)		(γ_1, γ_2) = (0,2)		(γ_1, γ_2) = (2,2)	
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
Our						
ABIAS^2	0.00003	0.00008	0.00170	0.00453	0.00039	0.00012
	(0.00003)	(0.00004)	(0.00019)	(0.00025)	(0.00015)	(0.00005)
AVAR	0.02844	0.01673	0.03455	0.02551	0.03283	0.02403
	(0.00143)	(0.00117)	(0.00139)	(0.00137)	(0.00125)	(0.00169)
AMSE	0.02847	0.01681	0.03625	0.03004	0.03322	0.02415
	(0.00143)	(0.00117)	(0.00144)	(0.00142)	(0.00124)	(0.00168)
M(h)	0.591	0.831	0.359	0.358	0.373	0.410
SD(h)	0.563	0.702	0.061	0.058	0.0955	0.152
A-C-Y (AFT, correct)						
ABIAS^2	0.00385	0.00374	0.00431	0.00488	0.00477	0.00486
	(0.00040)	(0.00036)	(0.00045)	(0.00045)	(0.00044)	(0.00044)
AVAR	0.01209	0.00866	0.01324	0.01015	0.01233	0.00960
	(0.00048)	(0.00047)	(0.00055)	(0.00050)	(0.00052)	(0.00047)
AMSE	0.01595	0.01240	0.01756	0.01503	0.01710	0.01447
	(0.00074)	(0.00074)	(0.00084)	(0.00084)	(0.00079)	(0.00081)
M(h)	1.899	1.649	1.852	1.669	1.889	1.680
SD(h)	1.105	0.977	1.134	0.965	1.088	0.970
A-C-Y(LOC, wrong)						
ABIAS^2	0.00202	0.00182	0.09768	0.20958	0.10368	0.02313
	(0.00028)	(0.00020)	(0.00143)	(0.00243)	(0.00117)	(0.00056)
AVAR	0.01153	0.00659	0.01424	0.01515	0.00955	0.00755
	(0.00047)	(0.00042)	(0.00037)	(0.00043)	(0.00042)	(0.00032)
AMSE	0.01356	0.00842	0.11193	0.22474	0.11323	0.03069
	(0.00064)	(0.00057)	(0.00132)	(0.00227)	(0.00116)	(0.00068)
M(h)	1.841	1.754	0.714	0.657	0.910	0.758
SD(h)	1.142	0.986	0.646	0.618	0.662	0.478

The number in the parenthesis below a quantity is the estimated standard deviation of that quantity
Calibration functions linear $\theta(z) = \exp(0.8z - 2)$, *quadratic* $\theta(z) = \exp(2 - 0.3(z-4)^2)$, *Our* proposed estimator (1), *A-C-Y (AFT)* Acar-Craiu-Yao estimator by fitting the correct marginal AFT models, *A-C-Y (LOC)* Acar-Craiu-Yao estimator by fitting wrong marginal LOC models, *M(h)* mean of the selected bandwidth, *SD(h)* standard deviation of the selected bandwidth. $n = 100$ and replications = 400

$$AVAR(\hat{\tau}) = \frac{1}{b-a}\int_a^b E\{\hat{\tau}(z) - E[\hat{\tau}(z)]\}^2 dz,$$

$$AMSE(\hat{\tau}) = \frac{1}{b-a}\int_a^b E\{[\hat{\tau}(z) - \tau(z)]^2\} dz = ABIAS^2(\hat{\tau}) + AVAR(\hat{\tau})$$

Table 2 Estimation of $\tau(z)$ without censoring and truncation with marginal AFT models under Frank copula (data structure 1)

	$(\gamma_1, \gamma_2) = (0,0)$		$(\gamma_1, \gamma_2) = (0,2)$		$(\gamma_1, \gamma_2) = (2,2)$	
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
Our						
ABIAS^2	0.00026	0.00020	0.00046	0.00081	0.00098	0.00035
	(0.00016)	(0.00093)	(0.00014)	(0.00015)	(0.00019)	(0.00011)
AVAR	0.03671	0.02731	0.04631	0.04203	0.04584	0.03662
	(0.00166)	(0.00134)	(0.00147)	(0.00177)	(0.00165)	(0.00161)
AMSE	0.03698	0.02752	0.04678	0.04284	0.04682	0.03697
	(0.00170)	(0.00136)	(0.00149)	(0.00180)	(0.00163)	(0.00159)
M(h)	0.622	0.762	0.360	0.350	0.354	0.376
SD(h)	0.547	0.681	0.095	0.048	0.056	0.086
A-C-Y						
ABIAS^2	0.01285	0.03907	0.01194	0.03861	0.01244	0.03994
	(0.00062)	(0.00122)	(0.00060)	(0.00113)	(0.00060)	(0.00107)
AVAR	0.01709	0.01828	0.01741	0.01696	0.01741	0.01635
	(0.00048)	(0.00060)	(0.00049)	(0.00059)	(0.00049)	(0.00057)
AMSE	0.02995	0.05736	0.02935	0.05557	0.02985	0.05630
	(0.00076)	(0.00109)	(0.00075)	(0.00103)	(0.00077)	(0.00098)
M(h)	0.686	1.555	0.706	1.573	0.743	1.602
SD(h)	0.657	1.129	0.729	1.089	0.746	1.085

The number in the parenthesis below a quantity is the estimated standard deviation of that quantity
Calibration functions linear $\theta(z) = \exp(0.8z - 2)$, *quadratic* $\theta(z) = \exp(2 - 0.3(z - 4)^2)$, *Our* proposed estimator (1) with incorrect Clayton copula, *A-C-Y* Acar-Craiu-Yao estimator by fitting the correct marginal AFT models with incorrect Clayton copula, *M(h)* mean of the selected bandwidth, *SD(h)* standard deviation of the selected bandwidth. $n = 100$ and replications = 400

respectively. Here $[a, b]$ denotes the support of covariate Z. That is, $a = 2$ and $b = 5$ for our simulation with $Z \sim Uniform(2, 5)$. Also note that we use the average quantities ABIAS^2, AVAR, and AMSE instead of the integrated quantities $\text{IBIAS}^2(\hat{\tau}) = \int_a^b \{E[\hat{\tau}(z)] - \tau(z)\}^2 dz$, $\text{IVAR}(\hat{\tau}) = \int_a^b E\{\hat{\tau}(z) - E[\hat{\tau}(z)]\}^2 dz$, and $\text{IMSE}(\hat{\tau}) = \int_a^b E\{[\hat{\tau}(z) - \tau(z)]^2\} dz$ in Acar et al. (2011). The integrated quantities differ from the corresponding averaged quantities by a factor of $b - a = 3$. We also observe that both estimators of $\tau(z)$ have larger variances for z closer to the boundary region.

Based on 400 simulation runs, the empirical accuracy measures ABIAS^2, AVAR, and AMSE are shown in Table 1. The standard deviations of ABIAS^2, AVAR, and AMSE are also reported. When fitting the correct AFT marginal model, Acar-Craiu-Yao estimator performs better. This is expected as Acar-Craiu-Yao estimator utilizes the correct likelihood function while our estimator only uses the pairwise concordance information without imposing any marginal model assumption. However, when the wrong LOC marginal model is used, ABIAS$^2(\hat{\tau})$ for the Acar-Craiu-Yao estimator

Table 3 Estimation of $\tau(z)$ based on bivariate data with independent censoring (data structure 2)

	$(\gamma_1, \gamma_2) = (0,0)$		$(\gamma_1, \gamma_2) = (0,2)$		$(\gamma_1, \gamma_2) = (2,2)$	
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
ABIAS^2	0.00020	0.00023	0.00039	0.00129	0.00096	0.00026
	(0.00009)	(0.00014)	(0.00012)	(0.00021)	(0.00029)	(0.00010)
AVAR	0.05766	0.02785	0.05706	0.03787	0.07536	0.04029
	(0.00338)	(0.00187)	(0.00237)	(0.00245)	(0.00344)	(0.00259)
AMSE	0.05787	0.02808	0.05746	0.03916	0.07632	0.04055
	(0.00337)	(0.00182)	(0.00237)	(0.00246)	(0.00349)	(0.00256)
M(h)	0.753	0.968	0.375	0.382	0.406	0.493
SD(h)	0.694	0.814	0.096	0.093	0.127	0.242
Cen1	0.249	0.249	0.001	0.001	0.246	0.249
Cen2	0.395	0.398	0.403	0.396	0.390	0.401

The number in the parenthesis below a quantity is the estimated standard deviation of that quantity
Calibration functions linear $\theta(z) = \exp(0.8z - 2)$, *quadratic* $\theta(z) = \exp(2 - 0.3(z-4)^2)$, *Cen1* censored rate of T_1, *Cen2* censored rate of T_2, *M(h)* mean of the selected bandwidth, *SD(h)* standard deviation of the selected bandwidth. $n = 100$ and replications = 400

Table 4 Estimation of $\tau(z)$ based on semi-competing risks data with independent censoring (data structure 3)

	$(\gamma_1, \gamma_2) = (0,0)$		$(\gamma_1, \gamma_2) = (0,2)$		$(\gamma_1, \gamma_2) = (2,2)$	
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
ABIAS^2	0.00108	0.00019	0.00070	0.00150	0.00052	0.00015
	(0.00050)	(0.00010)	(0.00020)	(0.00019)	(0.00019)	(0.00006)
AVAR	0.07918	0.03640	0.05651	0.03650	0.09390	0.04899
	(0.00421)	(0.00358)	(0.00212)	(0.00219)	(0.00428)	(0.00432)
AMSE	0.08027	0.03660	0.05722	0.03800	0.09442	0.04915
	(0.00442)	(0.00363)	(0.00211)	(0.00220)	(0.00435)	(0.00432)
M(h)	0.771	1.109	0.378	0.380	0.418	0.552
SD(h)	0.702	0.912	0.111	0.086	0.134	0.285
Cen1	0.391	0.352	0.002	0.001	0.392	0.349
Cen2	0.398	0.401	0.396	0.393	0.401	0.403

The number in the parenthesis below a quantity is the estimated standard deviation of that quantity
Calibration functions linear $\theta(z) = \exp(0.8z - 2)$, *quadratic* $\theta(z) = \exp(2 - 0.3(z-4)^2)$, *Cen1* censored rate of T_1, *Cen2* censored rate of T_2, *M(h)* mean of the selected bandwidth, *SD(h)* standard deviation of the selected bandwidth. $n = 100$ and replications = 400

inflates, resulting in larger AMSE than the proposed estimator. The results confirm that the proposed estimator is more robust since it does not require any assumption on the marginal models.

Since our simulations use the Acar-Craiu-Yao estimator with $(\hat{U}_{1,i}, \hat{U}_{2,i})$ estimated from the marginal model, the empirical accuracy measures reported here are larger than the numbers shown in Acar et al. (2011) where data of $(U_{1,i}, U_{2,i})$ are

Table 5 Estimation of $\tau(z)$ based on dependent truncated data with independent censoring (data structure 4)

	$(\gamma_1, \gamma_2) = (0,0)$		$(\gamma_1, \gamma_2) = (0,2)$		$(\gamma_1, \gamma_2) = (2,2)$	
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
ABIAS^2	0.00006	0.00005	0.00129	0.00340	0.02234	0.06050
	(0.00005)	(0.00005)	(0.00018)	(0.00024)	(0.00089)	(0.00143)
AVAR	0.03292	0.01709	0.03576	0.02410	0.04816	0.03878
	(0.00125)	(0.00095)	(0.00113)	(0.00093)	(0.00121)	(0.00127)
AMSE	0.03298	0.01715	0.03706	0.02750	0.07050	0.09929
	(0.00127)	(0.00098)	(0.00116)	(0.00100)	(0.00166)	(0.00222)
M(h)	0.609	0.915	0.363	0.359	0.336	0.333
SD(h)	0.525	0.773	0.071	0.064	0.019	0.008
Cen2	0.168	0.183	0.136	0.143	0.178	0.185

The number in the parenthesis below a quantity is the estimated standard deviation of that quantity
Calibration functions linear $\theta(z) = \exp(0.8z - 2)$, *quadratic* $\theta(z) = \exp(2 - 0.3(z-4)^2)$, *Cen2* censored rate of T_2, *M(h)* mean of the selected bandwidth, *SD(h)* standard deviation of the selected bandwidth. $n = 100$ and replications = 400

directly available. As expected, the extra estimation of $(U_{1,i}, U_{2,i})$ degrades the performance of the Acar-Craiu-Yao estimator but cannot be avoided in practical applications.

In Table 2, we examine the robustness issue for choosing the correct copula function for both our estimator and Acar-Craiu-Yao estimator. The simulation setting is similar to that of Table 1, but we set $(U_{1,i}, U_{2,i})$ to follow the Frank copula instead of the Clayton copula. Table 2 presents performance of our proposed estimator and Acar-Craiu-Yao estimator both assuming the incorrect Clayton copula. From the results, both estimators become less accurate with the wrongly specified copula. But the largest AMSE for our estimator under the wrong copula model is still $< 5\,\%$.

We also examine the performances of the proposed estimator when the data follow the second, third and forth data structures. Note that the Acar-Craiu-Yao procedure does not work for these situations and thus cannot be compared. In Table 3, the censoring variables (C_1, C_2) are generated independently from (T_1, T_2), where $C_1 = -1.5\exp(\gamma_2 Z)\log(U_3)$, $C_2 = -1.5\exp(\gamma_2 Z)\log(U_4)$, and U_3 and U_4 are uniform$(0, 1)$. In Table 4, C is generated independent from (T_1, T_2) and follows $C = -1.5\exp(\gamma_2 Z)\log(U_5)$, where U_5 follows uniform(0,1). In Table 5, $(U_{1,i}, U_{2,i})$ are generated from the extended Clayton copula $\tilde{C}_{\theta(Z_i)}(u_1, u_2)$ instead of $C_{\theta(Z_i)}(u_1, u_2)$, and C is generated independent from (T_1, T_2) and follows $C = -6\exp(\gamma_2 Z)\log(U_6)$, where U_6 follows uniform(0,1). Compared to the first data structure, the latter three data structures are subject to censoring. Thus the effective sample size decreases, resulting in larger estimation errors. In the worst case, AMSE is $< 10\,\%$.

5 Data example

For illustration, we apply the proposed methodology to analyze the bone marrow transplant (BMT) data on page 484 of Klein and Moeschberger (2003), which contains

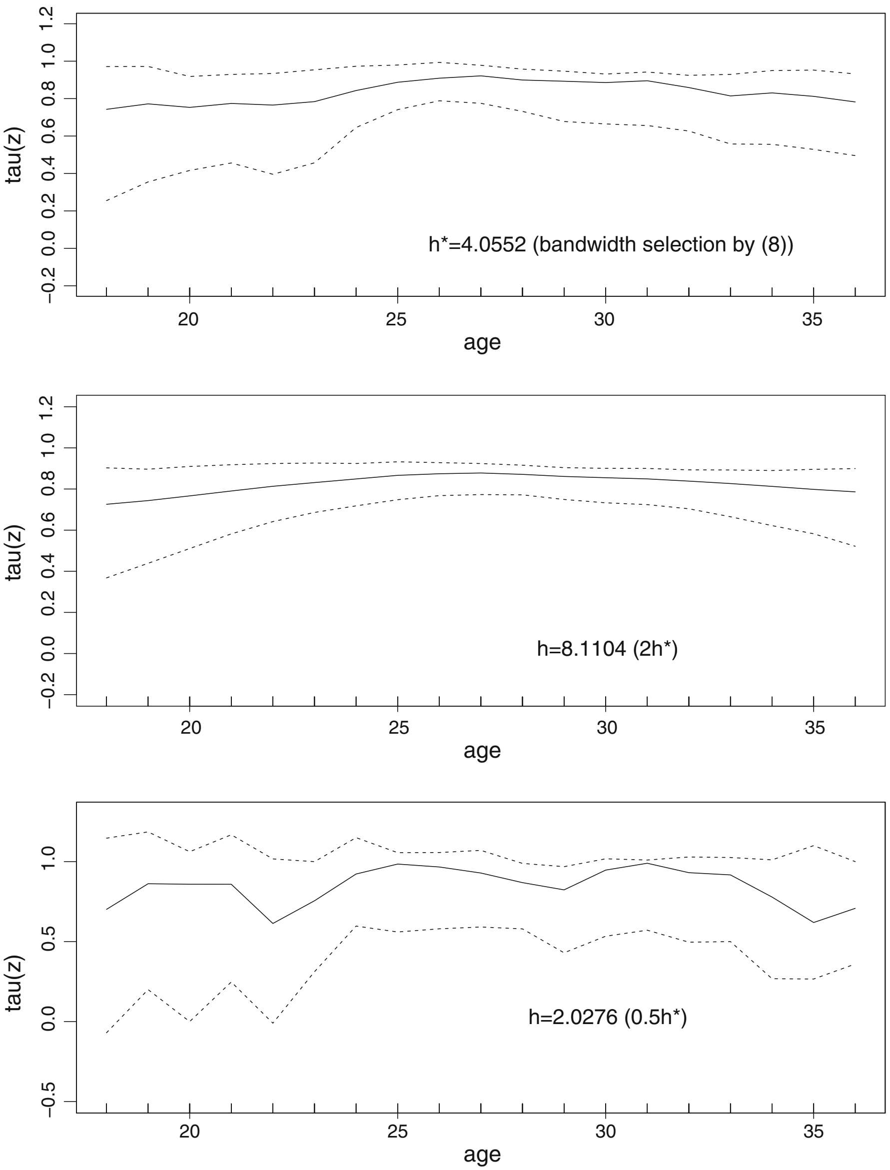

Fig. 1 The estimated function of $\tau(z)$ using Epanechnikov kernel for the bone marrow transplantation data where z corresponds to the age of the patients. The first plot is $\hat{\tau}(z)$ with the bandwidth h^* selected by (8), the second plot is $\hat{\tau}(z)$ with the bandwidth $2h^*$, the third plot is $\hat{\tau}(z)$ with the bandwidth $0.5h^*$

137 leukemia patients receiving bone marrow transplants. Of these patients, 40 died without relapse, 40 died after relapse, 54 were alive without relapse at the end of study period, 3 were alive after relapse at the end of study period. Using the time of transplantation as the origin, we consider the following two survival times: T_1 is the time to relapse of leukemia and T_2 is the time to death. Then T_1 and T_2 can be

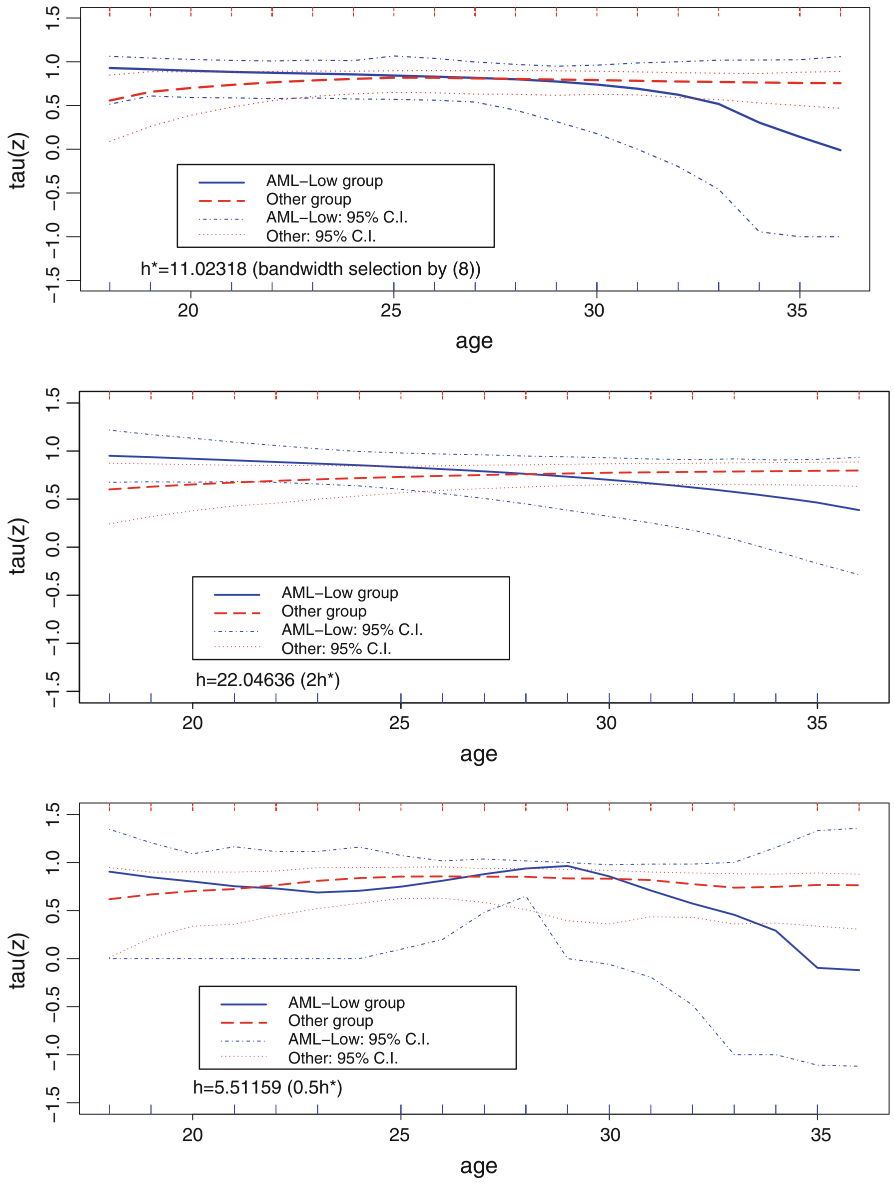

Fig. 2 The estimated function of $\tau(z)$ using Epanechnikov kernel for the bone marrow transplantation data: for the AML-Low group and the other patients group respectively where z corresponds to the age of the patients. The first plot is $\hat{\tau}(z)$ with the bandwidth h^* selected by (8), the second plot is $\hat{\tau}(z)$ with the bandwidth $2h^*$, the third plot is $\hat{\tau}(z)$ with the bandwidth $0.5h^*$

considered as the second structure of semi-competing risks data. Lakhal et al. (2008) showed that the two times T_1 and T_2 are correlated with Kendall's tau $\hat{\tau} = 0.80$.

Ding et al. (2009) considered marginal regression of T_1 and T_2 under dependent censoring. In particularly, based on their status at the time of transplantation, 54 patients

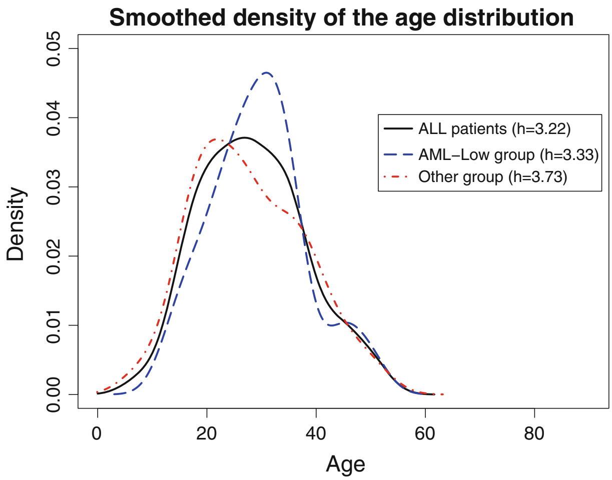

Fig. 3 The smoothed density for the age of patients in the bone marrow transplantation data. The density is estimated using the R-function density which uses a Gaussian kernel function and the automatically chosen bandwidth h values displayed in the figure

were classified into the AML-Low group. It is found that patients in the AML-Low group tend to have longer survival time (T_2) and relapse time T_1 than other patients in the study. The patient's age, a continuous covariate, does not have statistically significant effect on both marginal distributions of T_1 and T_2. However this analysis makes an implicit assumption that covariates (such as patient's age) do not influence the association between T_1 and T_2.

Here we illustrate our method by estimating the association between T_1 and T_2 as a function of patient's age (z). Note that the Acar-Craiu-Yao estimator is not applicable to this data structure. Figure 1 plots the proposed estimator of $\tau(z)$ for all patients. The two dotted lines are the 95 % pointwise confidence interval using the bootstrap percentile method. Specifically we generate bootstrap data from the original data and then compute $\hat{\tau}^*(z)$ based on the bootstrap sample. Repeating this procedure B times, we obtain $\hat{\tau}_b^*(z)$ ($b = 1, ..., B$). Then the $(1-\gamma)$ confidence interval of $\tau(z)$ can be constructed as $[\hat{\tau}^*_{(B\gamma/2)}(z), \hat{\tau}^*_{(B(1-\gamma/2))}(z)]$, where $\hat{\tau}^*_{(b)}(z)$ ($b = 1, ..., B$) are the order statistics of $\hat{\tau}_b^*(z)$ ($b = 1, ..., B$) and γ is the significance level. The result shows that the association between the relapse time and survival time is highly positive and significantly different from zero. It appears that patient's age does not significantly affect the association structure. We obtain that $\sum_{i=1}^n \hat{\tau}(z_i)/n$=0.81. The result agrees with the analysis by Lakhal et al. (2008) and Ding et al. (2009).

We can take a closer look at the data by applying the proposed analysis to different risk groups. In Fig. 2, we plot the estimated function of $\tau(z)$ for the AML-Low group and the group of other patients. The plot shows that the association $\tau(z)$ decreases with patient's age in the AML-Low group, and increases with patient's age in the other group. Although the difference in the association pattern is not significant due to small number of older patients in the two groups, our analysis still reveals some interesting phenomenon to medical practitioners. In the first two figures, we also plot the results using half and double the bandwidth selected by (8). The pattern of age effect on association remains the same. However these additional two bandwidths do seem to undersmooth and oversmooth the association patterns.

Note that Figs. 1 and 2 only draw the middle range of the age distribution where there are enough patients for estimating $\tau(z)$. We do not include the results for few very young or very old patients because the resulting estimates of $\tau(z)$ have high variance and hence are not reliable for these z values. Figure 3 shows the estimated density of the age distribution in this data set.

6 Concluding remarks

We propose a nonparametric approach to estimating the association parameter as a function of covariates. This novel technique uses the pairwise concordance indicator as the response variable and so is not a direct application of existing local linear regression methods. Most traditional nonparametric methods are rank-based procedures which cannot be immediately applied when censoring or truncation occurs. As a result, the rank statistics are often re-expressed in terms of pairwise order relationship. The proposed approach is the first smoothing technique based on pairwise quantities.

When complete data are available, our method can be compared with the likelihood-based approach of Acar et al. (2011). Although the proposed method seems to be less efficient, it does not require specifying the marginal models and hence is more robust. Furthermore our approach, like the Acar-Craiu-Yao approach, also works for non-Clayton copula models. For those models with $\alpha(z) = 2\int\int C_{\theta(z)}(u, v)dC_{\theta(z)}(u, v)$, we can also translate $\alpha(z)$ into the parameter $\theta(z)$ of the given copula.

The Clayton assumption is not needed for complete data since the proposed method directly estimates $\alpha(z) = \Pr(\delta_{ij} = 1|Z_i = Z_j = z)$ which involves no model assumption. For the other three data structures the Clayton model is assumed to result in non-informative missing patterns for δ_{ij}. With the non-informative missing patterns, $\alpha(z)$ is the same as the concordance probability among comparable pairs $\alpha^*(z)$ which is estimated by our local linear estimator. Extension to non-Clayton copula families without the non-informative missing property would produce additional problems which are not our focus here.

Appendix 1: Proof for Equation (2)

To show that the missing mechanism is non-informative under censoring, we consider the localized association measure in Oakes (1989). Let $\alpha^*(s, t; z) = Pr(\delta_{ij} = 1|T_{1,i} \wedge T_{1,j} = s, T_{2,i} \wedge T_{2,j} = t, Z_i = Z_j = z)$. For Clayton copula, this localized measure remains a constant over time, that is,

$$\alpha^*(s, t; z) = \alpha(z). \tag{9}$$

This property ensures that Eq. (2)holds under censoring. Let $f(s, t; z)$ denote the joint density function for $T_{1,i} \wedge T_{1,j}$ and $T_{2,i} \wedge T_{2,j}$ conditional on $Z_i = Z_j = z$. Then $\alpha(z) = Pr(\delta_{ij} = 1|Z_i = Z_j = z) = \int_{s=0}^{\infty}\int_{t=0}^{\infty} f(s, t; z)\alpha^*(s, t; z)dsdt$.

For the second data structure, let $G(s,t;z) = Pr(Y_{1,i} \wedge Y_{1,j} > s, Y_{2,i} \wedge Y_{2,j} > t | Z_i = Z_j = z)$ denote the joint conditional survival function for pairwise minimum of the censoring times. Then

$$Pr(\delta_{ij} = 1 | \Delta_{ij} = 1, Z_i = Z_j = z) = \frac{\int_{s=0}^{\infty} \int_{t=0}^{\infty} f(s,t;z) G(s,t;z) \alpha^*(s,t;z) dsdt}{\int_{s=0}^{\infty} \int_{t=0}^{\infty} f(s,t;z) G(s,t;z) dsdt},$$

which equals $\alpha(z)$ by (9).

For the third data structure, let $G^*(t;z) = Pr(Y_i \wedge Y_j > t | Z_i = Z_j = z)$ denote the conditional survival function for $Y_i \wedge Y_j$. Then

$$Pr(\delta_{ij} = 1 | \Delta_{ij} = 1, Z_i = Z_j = z) = \frac{\int_{s=0}^{\infty} \int_{t=s}^{\infty} f(s,t;z) G^*(t;z) \alpha^*(s,t;z) dsdt}{\int_{s=0}^{\infty} \int_{t=s}^{\infty} f(s,t;z) G^*(t;z) dsdt},$$

which again equals $\alpha(z)$ by (9).

For the fourth data structure, consider the localized measure $\alpha^{**}(s,t;z) = Pr(\delta_{ij} = 1 | T_{1,i} \vee T_{1,j} = s, T_{2,i} \wedge T_{2,j} = t, Z_i = Z_j = z)$. Then $\alpha^{**}(s,t;z)$ remains a constant $\alpha(z)$ over time under the pseudo-Clayton copula due to the correspondence to $\alpha^*(s,t;z)$ for the Clayton copula. Again, let $G^*(t;z) = Pr(Y_i \wedge Y_j > t | Z_i = Z_j = z)$ denote the conditional survival function for $Y_i \wedge Y_j$. Then

$$Pr(\delta_{ij} = 1 | \Delta_{ij} = 1, Z_i = Z_j = z) = \frac{\int_{s=0}^{\infty} \int_{t=s}^{\infty} f(s,t;z) G^*(t;z) \alpha^{**}(s,t;z) dsdt}{\int_{s=0}^{\infty} \int_{t=s}^{\infty} f(s,t;z) G^*(t;z) dsdt}$$

which equals $\alpha(z)$ by constancy of $\alpha^{**}(s,t;z)$ over time.

Appendix 2: Proof of the Theorem 1

First we rewrite the solution to (1) in matrix terms. The design matrix $\mathbf{Z}_z$ for (1) is denoted by

$$\mathbf{Z}_z = (1, (Z_{ij} - \mathbf{z})^T)_{i<j} = \begin{pmatrix} 1 & (Z_1 - z)^T & (Z_2 - z)^T \\ 1 & (Z_1 - z)^T & (Z_3 - z)^T \\ \vdots & \vdots & \vdots \\ 1 & (Z_{n-1} - z)^T & (Z_n - z)^T \end{pmatrix}.$$

The weight matrix is denoted by $\mathbf{W}_z = diag\{\Delta_{ij} K_H(Z_{ij} - \mathbf{z})\}_{i<j}$. And the local linear estimator becomes

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (\mathbf{Z}_z^T \mathbf{W}_z \mathbf{Z}_z)^{-1} \mathbf{Z}_z^T \mathbf{W}_z \boldsymbol{\delta}.$$

Thus

$$\hat{\alpha} = e_1^T(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^T\mathbf{W}_z\boldsymbol{\delta}, \tag{10}$$

where e_1 is the $(2d+1)\times 1$ vector with 1 in the first entry and all other entries 0.

To simplify the notations, in the following calculations, we denote $E[\hat{\alpha}]$ and $Var[\hat{\alpha}]$ as the mean and variance conditional on $Z_1, ..., Z_n$ (or equivalently conditional on $\mathbf{Z}_z$). That is, computing the mean and variance coming from randomness of $\boldsymbol{\delta}$ only.

Denote the gradient of $\alpha(z)$ as $\mathcal{D}_\alpha(z) = \begin{pmatrix} \frac{\partial}{\partial z_1}\tilde{\alpha}(z_1, z_2) \\ \frac{\partial}{\partial z_2}\tilde{\alpha}(z_1, z_2) \end{pmatrix}_{z_1=z_2=z}$ and the Hessian matrix of $\alpha(z)$ as

$$\mathcal{H}_\alpha(z) = \begin{pmatrix} \frac{\partial^2}{\partial z_1 \partial z_1^T}\tilde{\alpha}(z_1, z_2) & \frac{\partial^2}{\partial z_1 \partial z_2^T}\tilde{\alpha}(z_1, z_2) \\ \frac{\partial^2}{\partial z_2 \partial z_1^T}\tilde{\alpha}(z_1, z_2) & \frac{\partial^2}{\partial z_2 \partial z_2^T}\tilde{\alpha}(z_1, z_2) \end{pmatrix}_{z_1=z_2=z}.$$

Then by the Taylor expansion,

$$\tilde{\alpha}(Z_{ij}) = \alpha(z) + (Z_{ij}-\mathbf{z})^T\mathcal{D}_\alpha(z) + \frac{1}{2}(Z_{ij}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{ij}-\mathbf{z}) + o_p(\|Z_{ij}-\mathbf{z}\|^2)$$

Denote $\mathbf{A} = E(\boldsymbol{\delta}|Z_1, ..., Z_n) = (\tilde{\alpha}(Z_{ij}))_{i<j}$. Hence we have

$$\mathbf{A} = \mathbf{Z}_z\begin{pmatrix} \alpha(z) \\ \mathcal{D}_\alpha(z) \end{pmatrix} + \frac{1}{2}\mathbf{Q}_\alpha(z) + \mathbf{R}_\alpha(z)$$

where

$$\mathbf{Q}_\alpha(z) = \left((Z_{ij}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{ij}-\mathbf{z})\right)_{i<j}$$

$$= \begin{pmatrix} (Z_{12}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{12}-\mathbf{z}) \\ (Z_{13}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{13}-\mathbf{z}) \\ \vdots \\ (Z_{1n}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{1n}-\mathbf{z}) \\ (Z_{23}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{23}-\mathbf{z}) \\ \vdots \\ (Z_{(n-1)n}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{(n-1)n}-\mathbf{z}) \end{pmatrix},$$

and $\mathbf{R}_\alpha(z)$ is the vector of Taylor series remainder terms.

Accordingly, plug the above expression into (10), we get

$$\begin{aligned} E(\hat{\alpha}|Z_1, ..., Z_n) &= e_1^T(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^T\mathbf{W}_z\mathbf{A} \\ &= \alpha(z) + e_1^T(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^T\mathbf{W}_z[\tfrac{1}{2}\mathbf{Q}_\alpha(z) + \mathbf{R}_\alpha(z)]. \end{aligned}$$

As in the usual local linear regression setting, the bias of $\hat{\alpha}$ is about

$$\frac{1}{2}e_1^T(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Q}_\alpha(z) \tag{11}$$

because the term $e_1^T(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^T\mathbf{W}_z\mathbf{R}_\alpha(z)$ is of smaller order. The variance of $\hat{\alpha}$ is

$$e_1^T(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^T\mathbf{W}_z Var(\boldsymbol{\delta})\mathbf{W}_z\mathbf{Z}_z(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}e_1. \tag{12}$$

However, unlike the usual local linear regression, these expressions need more careful analysis because (a) the terms δ_{ij} in $\boldsymbol{\delta}$ are not independent and (b) the weight matrix involves extra correlated random variables Δ_{ij}.

First, from Appendix 3 we obtain

$$\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z = \frac{n(n-1)}{2}\tilde{g}(z,z)[f_Z(z)]^2\tilde{H}[\begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K)I_{2d}\end{pmatrix} + o_p(1)]\tilde{H}, \tag{13}$$

where

$$\tilde{H} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & H & 0 \\ 0 & 0 & H \end{pmatrix}. \tag{14}$$

Then from Appendix 4, We have

$$\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Q}_\alpha(z) = \begin{pmatrix} n(n-1)\tilde{g}(z,z)[f_Z(z)]^2\mu_2(K)tr\{H^2\mathcal{H}_\alpha(z)\} + o_p(n^2 tr(H^2)) \\ O_p\left[n^2\begin{pmatrix} H^3\mathbf{1} \\ H^3\mathbf{1}\end{pmatrix}\right] \end{pmatrix}.$$

Thus the bias of $\hat{\alpha}$ is

$$\frac{1}{2}e_1^T(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Q}_\alpha(z) + o_p(tr(H^2)) = \mu_2(K)tr\{H^2\mathcal{H}_\alpha(z)\} + o_p(tr(H^2))$$

Secondly from Appendix 5,

$$\mathbf{Z}_z^T\mathbf{W}_z Var(\boldsymbol{\delta})\mathbf{W}_z\mathbf{Z}_z \to n(n-1)(n-2)|H|^{-1}\breve{g}(z,z,z)\breve{\alpha}(z,z,z)\mu_0(K^2)[f_Z(z)]^3 \times \begin{pmatrix} 1 & O_p[H\mathbf{1}]^T & O_p[H\mathbf{1}]^T \\ O_p[H\mathbf{1}] & O_p[H^2] & O_p[H^2] \\ O_p[H\mathbf{1}] & O_p[H^2] & O_p[H^2] \end{pmatrix}.$$

Using this and (13), we have the variance of $\hat{\alpha}$ as

$$
\begin{aligned}
& e_1^T(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}\mathbf{Z}_z^T\mathbf{W}_z Var(\boldsymbol{\delta})\mathbf{W}_z\mathbf{Z}_z(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)^{-1}e_1 \\
&= \frac{4(n-2)\check{g}(z,z,z)\check{\alpha}(z,z,z)\mu_0(K^2)}{n(n-1)|H|[\tilde{g}(z,z)]^2 f_Z(z)} + o_p(n^{-1}|H|^{-1}).
\end{aligned}
$$

Notice that the order of variance is $O_p(n^{-1}|H|^{-1})$ instead of $O_p(n^{-2}|H|^{-1})$, the latter of which is the variance when δ_{ij}'s are independent for usual local linear regression setting.

Appendix 3: Analysis of the matrix $(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)$

The matrix $(\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z)$ can be rewritten as

$$
\mathbf{Z}_z^T\mathbf{W}_z\mathbf{Z}_z = \begin{pmatrix} S_0 & S_1^T \\ S_1 & S_2 \end{pmatrix},
$$

where

$$
\begin{aligned}
S_0 &= \sum\nolimits_{i<j}\Delta_{ij}K_H(Z_{ij}-\mathbf{z}), \\
S_1 &= \sum\nolimits_{i<j}\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z}), \\
S_2 &= \sum\nolimits_{i<j}\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z})^T.
\end{aligned}
$$

Compared to ordinary local linear regression, the terms $S_k, k=0,1,2$ here (a) are not i.i.d. sums and (b) contains the extra random variables Δ_{ij}. Thus more careful asymptotic analysis is required. Denote

$$
\begin{aligned}
\tilde{g}(z_1,z_2) &= E[\Delta_{12}|Z_1=z_1, Z_2=z_2], \\
\check{g}(z_1,z_2,z_3) &= E[\Delta_{12}\Delta_{13}|Z_1=z_1, Z_2=z_2, Z_3=z_3].
\end{aligned}
$$

Then we have

$$
\begin{aligned}
E\left[S_0|Z_1,...,Z_n\right] &= E\left[\sum\nolimits_{i<j}\Delta_{ij}K_H(Z_{ij}-\mathbf{z})|Z_1,...,Z_n\right] \\
&= \sum\nolimits_{i<j}\tilde{g}(Z_i,Z_j)K_H(Z_{ij}-\mathbf{z}) \\
&= \frac{1}{2}\sum\nolimits_{i\neq j}\tilde{g}(Z_i,Z_j)K_H(Z_{ij}-\mathbf{z}).
\end{aligned}
$$

The above expression converges to the limit

$$
\begin{aligned}
\frac{1}{2}\sum\nolimits_{i\neq j}\tilde{g}(Z_i,Z_j)K_H(Z_{ij}-\mathbf{z}) \to \frac{n(n-1)}{2}\int\int \tilde{g}(z_1,z_2)|H|^{-2}K(H^{-1}(z_1-z)) \\
\times K(H^{-1}(z_2-z))f_Z(z_1)f_Z(z_2)dz_1dz_2
\end{aligned}
$$

$$= \frac{n(n-1)}{2} \int\int \tilde{g}(z + Hu_1, z + Hu_2) K(u_1) K(u_2) f_Z(z + Hu_1)$$
$$\times\ f_Z(z + u_2) du_1 du_2 = \frac{n(n-1)}{2} \tilde{g}(z, z) [f_Z(z)]^2 [1 + o(1)],$$

using change of variable $u_1 = H^{-1}(z_1 - z)$ and $u_2 = H^{-1}(z_2 - z)$. Hence $E[S_0|Z_1, ..., Z_n] = \frac{n(n-1)}{2} \tilde{g}(z, z)[f_Z(z)]^2[1 + o(1)]$.

The variance of S_0 is

$$\begin{aligned} Var[S_0|Z_1, ..., Z_n] &= \sum_{i<j} \sum_{k<l} Cov[\Delta_{ij} K_H(Z_{ij} - \mathbf{z}), \Delta_{kl} K_H(Z_{kl} - \mathbf{z})|Z_1, ..., Z_n] \\ &= \sum_{i<j} Var[\Delta_{ij} K_H(Z_{ij} - \mathbf{z})] \\ &\quad + \sum_{i<(j\neq k)} Cov[\Delta_{ij} K_H(Z_{ij} - \mathbf{z}), \Delta_{ik} K_H(Z_{ik} - \mathbf{z})] \\ &\quad + \sum_{(i\neq k)<j} Cov[\Delta_{ij} K_H(Z_{ij} - \mathbf{z}), \Delta_{kj} K_H(Z_{kj} - \mathbf{z})] \\ &\quad + \sum_{i<j<k} Cov[\Delta_{ij} K_H(Z_{ij} - \mathbf{z}), \Delta_{jk} K_H(Z_{jk} - \mathbf{z})] \\ &\quad + \sum_{k<i<j} Cov[\Delta_{ij} K_H(Z_{ij} - \mathbf{z}), \Delta_{ki} K_H(Z_{ki} - \mathbf{z})] + 0 \\ &= A_1 + A_2 + A_3 + A_4 + A_5. \end{aligned}$$

Here the Var and Cov in A_i's are conditional on $Z_1, ..., Z_n$ as before. We abuse the notations for abbreviation.

By symmetry, the first term

$$A_1 = \sum_{i<j} Var\left[\Delta_{ij} K_H(Z_{ij} - \mathbf{z})\right] = \frac{n(n-1)}{2} Var\left[\Delta_{12} K_H(Z_{12} - \mathbf{z})\right],$$

which becomes

$$\frac{n(n-1)}{2} E\left\{[\Delta_{12} K_H(Z_{12} - \mathbf{z})]^2\right\} - \frac{n(n-1)}{2} \{E\left[\Delta_{12} K_H(Z_{12} - \mathbf{z})\right]\}^2.$$

Hence

$$\begin{aligned} A_1 &= \frac{n(n-1)}{2} E\{\tilde{g}(Z_1, Z_2)[K_H(Z_{12} - \mathbf{z})]^2\} - \frac{n(n-1)}{2} \\ &\quad \times \{E[\tilde{g}(Z_1, Z_2) K_H(Z_{12} - \mathbf{z})]\}^2 \\ &= \frac{n(n-1)}{2} \int\int \tilde{g}(z_1, z_2)[|H|^{-2} K(H^{-1}(z_1 - \mathbf{z})) K(H^{-1}(z_2 - \mathbf{z}))]^2 \end{aligned}$$

$$
\begin{aligned}
&\times\ f_Z(z_1)f_Z(z_2)dz_1dz_2 - \frac{n(n-1)}{2}\left[\int\int \tilde{g}(z_1,z_2)|H|^{-2}K(H^{-1}(z_1-\mathbf{z}))\right.\\
&\times\ \left. K(H^{-1}(z_2-\mathbf{z}))f_Z(z_1)f_Z(z_2)dz_1dz_2\right]^2\\
=\ &\frac{n(n-1)}{2}|H|^{-2}\int\int \tilde{g}(z+Hu_1,z+Hu_2)[K(u_1)K(u_2)]^2\\
&\times\ f_Z(z+Hu_1)f_Z(z+Hu_2)du_1du_2\\
&-\frac{n(n-1)}{2}\left\{\int\int \tilde{g}(z+Hu_1,z+Hu_2)K(u_1)K(u_2)\right.\\
&\times\ \left. f_Z(z+Hu_1)f_Z(z+Hu_2)du_1du_2\right\}^2 \to \frac{n(n-1)}{2}|H|^{-2}\\
&\times\ \tilde{g}(z,z)[\mu_0(K^2)]^2[f_Z(z)]^2 - \frac{n(n-1)}{2}\{[\mu_1(K)]^2[\tilde{g}(z,z)f_Z(z)]^2\}^2\\
\to\ &\frac{n(n-1)}{2}|H|^{-2}\tilde{g}(z,z)[\mu_0(K^2)]^2[f_Z(z)]^2, \qquad (15)
\end{aligned}
$$

where $\mu_0(K^2) = \int[K(u)]^2du$ and $\mu_1(K) = \int K(u)udu$. And in the last step we drop a term $\frac{n(n-1)}{2}\{E[\Delta_{12}K_H(Z_{12}-\mathbf{z})]\}^2 = O_p(n^2)$ since $|H| \to 0$. The second term becomes

$$
\begin{aligned}
A_2\ =\ &\sum_{i<(j\neq k)} Cov[\Delta_{ij}K_H(Z_{ij}-\mathbf{z}),\Delta_{ik}K_H(Z_{ik}-\mathbf{z})]\\
=\ &\frac{n(n-1)(n-2)}{4}Cov[\Delta_{12}K_H(Z_{12}-\mathbf{z}),\Delta_{13}K_H(Z_{13}-\mathbf{z})]\\
=\ &\frac{n(n-1)(n-2)}{4}E[\Delta_{12}\Delta_{13}K_H(Z_{12}-\mathbf{z})K_H(Z_{13}-\mathbf{z})]\\
&-\ \frac{n(n-1)(n-2)}{4}\{E[\Delta_{12}K_H(Z_{12}-\mathbf{z})]\}^2\\
=\ &\frac{n(n-1)(n-2)}{4}E[\check{g}(Z_1,Z_2,Z_3)K_H(Z_{12}-\mathbf{z})K_H(Z_{13}-\mathbf{z})] - O_p(n^3)\\
=\ &\frac{n(n-1)(n-2)}{4}\int\int\int \check{g}(z_1,z_2,z_3)[|H|^{-2}K(H^{-1}(z_1-\mathbf{z}))\\
&\times\ K(H^{-1}(z_2-\mathbf{z}))][|H|^{-2}K(H^{-1}(z_1-\mathbf{z}))K(H^{-1}(z_3-\mathbf{z}))]f_Z(z_1)f_Z(z_2)\\
&\times\ f_Z(z_3)dz_1dz_2dz_3 - O_p(n^3)\\
\to\ &\frac{n(n-1)(n-2)}{4}|H|^{-1}\check{g}(z,z,z)\mu_0(K^2)[f_Z(z)]^3. \qquad (16)
\end{aligned}
$$

Then by symmetry, it is easy to see that

$$
A_2 = A_3 = A_4 = A_5 \to \frac{n(n-1)(n-2)}{4}|H|^{-1}\check{g}(z,z,z)\mu_0(K^2)[f_Z(z)]^3.
$$

Thus

$$
\begin{aligned}
Var[S_0|Z_1, ..., Z_n] &\to \frac{n(n-1)}{2}|H|^{-2}\tilde{g}(z,z)[\mu_0(K^2)]^2 \\
&\quad \times [f_Z(z)]^2 + n(n-1)(n-2)|H|^{-1}\breve{g}(z,z,z)\mu_0(K^2)[f_Z(z)]^3 \\
&= O_p(n^2|H|^{-2} + n^3|H|^{-1}) \\
&= o_p(n^4) = o_p\{[E(S_0|Z_1, ..., Z_n)]^2\},
\end{aligned}
$$

since $n|H| \to \infty$ as assumed in condition (ii). Therefore $S_0 \to \frac{n(n-1)}{2}\tilde{g}(z,z)[f_Z(z)]^2$. Notice that if δ_{ij}'s are $i.i.d$, the variance of S_0 would be of order $O_p(n^2|H|^{-2}$ which is dominated by the extra term $O_p(n^3|H|^{-1})$ due to the correlation among some pairs of δ_{ij}'s. Fortunately, the new order $O_p(n^3|H|^{-1})$ is still dominated by the order of $[E(S_0|Z_1, ..., Z_n)]^2 = O_p(n^4)$, ensuring the convergence of S_0.

Similarly, it follows that

$$
\begin{aligned}
&E[S_1|Z_1, ..., Z_n] \\
&\quad = \frac{n(n-1)}{2}E[\tilde{g}(Z_1, Z_2)K_H(Z_{12} - \mathbf{z})(Z_{12} - \mathbf{z})] \\
&\quad = \frac{n(n-1)}{2}\int\int \tilde{g}(z_1, z_2)|H|^{-2}K(H^{-1}(z_1 - z))K(H^{-1}(z_2 - z)) \\
&\qquad \times\ f_Z(z_1)f_Z(z_2)\begin{pmatrix} z_1 - z \\ z_2 - z \end{pmatrix} dz_1dz_2 \\
&\quad = \frac{n(n-1)}{2}\int\int K(u_1)K(u_2)\tilde{g}(z + Hu_1, z + Hu_2)f_Z(z + Hu_1) \\
&\qquad \times\ f_Z(z + Hu_2)H\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} du_1du_2 \\
&\quad = \frac{n(n-1)}{2}H\tilde{g}(z,z)[f_Z(z)]^2\begin{pmatrix} \mu_0(K)\mu_1(K) \\ \mu_0(K)\mu_1(K) \end{pmatrix} + o_p(n^2|H|).
\end{aligned}
\tag{17}
$$

Recall that $\mu_0(K) = \int K(u)du = 1$ is a scalar, $\mu_1(K) = \int K(u)udu$ is a d-dimensional vector. With the technical assumption of kernel K, $\mu_1(K)$ is a zero vector. So $E[S_1|Z_1, ..., Z_n] = o_p(n^2|H|)$. If we look more carefully,

$$
\begin{aligned}
&\tilde{g}(z + Hu_1, z + Hu_2)f_Z(z + Hu_1)f_Z(z + Hu_2) \\
&\quad = \tilde{g}(z,z)[f_Z(z)]^2 + f_Z(z)D^T_{gf_Z}(z)Hu_1 + f_Z(z)D^T_{gf_Z}(z)Hu_2 + O_p(|H|^2)
\end{aligned}
$$

where $D_{gf_Z}(z) = \{\frac{\partial}{\partial x}[\tilde{g}(x,y)f_Z(x)]\}_{x=y=z}$ is a d-dimensional vector. By symmetry of $\tilde{g}(x,y)$, we also have $D_{gf_Z}(z) = \{\frac{\partial}{\partial y}[\tilde{g}(x,y)f_Z(y)]\}_{x=y=z}$. Using this expansion,

$$
\begin{aligned}
E[S_1|Z_1, ..., Z_n] &= \frac{n(n-1)}{2}\mu_2(K)\begin{pmatrix} H^2D_{gf_Z}(z) \\ H^2D_{gf_Z}(z) \end{pmatrix} f_Z(z)[1 + o(1)] \\
&= O_p(n^2|H|^2) = o_p(n^2|H|).
\end{aligned}
$$

The variance of S_1 follows that

$$
\begin{aligned}
Var[S_1|Z_1,...,Z_n] &= \sum_{i<j}\sum_{k<l} Cov[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z}),\\
&\quad \times \Delta_{kl}K_H(Z_{kl}-\mathbf{z})(Z_{kl}-\mathbf{z})]\\
&= \sum_{i<j} Var[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z})]\\
&\quad + \sum_{i<(j\neq k)} Cov[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z}), \Delta_{ik}K_H(Z_{ik}-\mathbf{z})(Z_{ik}-\mathbf{z})]\\
&\quad + \sum_{(i\neq k)<j} Cov[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z}), \Delta_{kj}K_H(Z_{kj}-\mathbf{z})(Z_{kj}-\mathbf{z})]\\
&\quad + \sum_{i<j<k} Cov[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z}), \Delta_{jk}K_H(Z_{jk}-\mathbf{z})(Z_{jk}-\mathbf{z})]\\
&\quad + \sum_{k<i<j} Cov[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z}), \Delta_{ki}K_H(Z_{ki}-\mathbf{z})(Z_{ki}-\mathbf{z})] + 0\\
&= O_p(n^2) + O_p(n^3|H|) + O_p(n^3|H|) + O_p(n^3|H|) + O_p(n^3|H|)\\
&= O_p(n^2) + O_p(n^3|H|)\\
&= o_p(n^4|H|^2),
\end{aligned}
$$

with the last step coming from condition $n|H| \to \infty$. So we get $S_1 = o_p(n^2|H|)$.

$$
\begin{aligned}
E[S_2|Z_1,...,Z_n] &= \frac{n(n-1)}{2} E[\tilde{g}(Z_1,Z_2)K_H(Z_{12}-\mathbf{z})(Z_{12}-\mathbf{z})(Z_{12}-\mathbf{z})^T]\\
&= \frac{n(n-1)}{2}\int\int \tilde{g}(z_1,z_2)|H|^{-2}K(H^{-1}(z_1-z))K(H^{-1}(z_2-z))\\
&\quad \times\ f_Z(z_1)f_Z(z_2)\begin{pmatrix}(z_1-z)(z_1-z)^T & (z_1-z)(z_2-z)^T\\ (z_2-z)(z_1-z)^T & (z_2-z)(z_2-z)^T\end{pmatrix}dz_1dz_2\\
&= \frac{n(n-1)}{2}\int\int K(u_1)K(u_2)\tilde{g}(z+Hu_1, z+Hu_2)\\
&\quad \times\ f_Z(z+Hu_1)f_Z(z+Hu_2)\begin{pmatrix}Hu_1u_1^TH & Hu_1u_2^TH\\ Hu_2u_1^TH & Hu_2u_2^TH\end{pmatrix}du_1du_2\\
&= \frac{n(n-1)}{2}\tilde{g}(z,z)[f_Z(z)]^2\mu_2(K)\begin{pmatrix}H^2 & 0\\ 0 & H^2\end{pmatrix} + o_p(n^2|H|^2),
\end{aligned}
\tag{18}
$$

and

$$
Var(S_2|Z_1,...,Z_n) = O_p(n^2|H|^2) + O_p(n^3|H|^3) = o_p(n^4|H|^4).
$$

Hence,

$$S_2 = \mu_2(K) \begin{pmatrix} H^2 & 0 \\ 0 & H^2 \end{pmatrix} + o_p(n^2|H|^2).$$

Combining all the results together, we get Eq. (13)

$$\mathbf{Z}_z^T \mathbf{W}_z \mathbf{Z}_z = \frac{n(n-1)}{2} \tilde{g}(z,z)[f_Z(z)]^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mu_2(K)H^2 & 0 \\ 0 & 0 & \mu_2(K)H^2 \end{pmatrix} [1 + o_p(1)].$$

Appendix 4: Analysis of the vector $\mathbf{Z}_z^T \mathbf{W}_z \mathbf{Q}_\alpha(z)$

Now we have

$$\mathbf{Z}_z^T \mathbf{W}_z \mathbf{Q}_\alpha(z) = \begin{pmatrix} \sum_{i<j}[\Delta_{ij} K_H(Z_{ij} - \mathbf{z})(Z_{ij} - \mathbf{z})^T \mathcal{H}_\alpha(z)(Z_{ij} - \mathbf{z})] \\ \sum_{i<j}[\Delta_{ij} K_H(Z_{ij} - \mathbf{z})(Z_{ij} - \mathbf{z})^T \mathcal{H}_\alpha(z)(Z_{ij} - \mathbf{z})](Z_{ij} - \mathbf{z}) \end{pmatrix}.$$

First,

$$\begin{aligned}
& E\left\{\sum_{i<j}[\Delta_{ij} K_H(Z_{ij} - \mathbf{z})(Z_{ij} - \mathbf{z})^T \mathcal{H}_\alpha(z)(Z_{ij} - \mathbf{z})]\right\} \\
&= \frac{n(n-1)}{2} E[\tilde{g}(Z_1, Z_2) K_H(Z_{12} - \mathbf{z})(Z_{12} - \mathbf{z})^T \mathcal{H}_\alpha(z)(Z_{12} - \mathbf{z})] \\
&= \frac{n(n-1)}{2} \int\int \tilde{g}(z_1, z_2)|H|^{-2} K(H^{-1}(z_1 - z)) K(H^{-1}(z_2 - z)) \\
&\quad \times \; f_Z(z_1) f_Z(z_2)((z_1 - z)^T, (z_2 - z)^T)\mathcal{H}_\alpha(z) \begin{pmatrix} z_1 - z \\ z_2 - z \end{pmatrix} dz_1 dz_2 \qquad (19) \\
&= \frac{n(n-1)}{2} \int\int K(u_1)K(u_2)\tilde{g}(z + Hu_1, z + Hu_2) \\
&\quad \times \; f_Z(z + Hu_1) f_Z(z + Hu_2)[u_1^T H \mathcal{H}_\alpha(z) H u_1 + u_2^T H \mathcal{H}_\alpha(z) H u_2] du_1 du_2 \\
&= \frac{n(n-1)}{2} \int\int K(u_1)K(u_2)\tilde{g}(z + Hu_1, z + Hu_2) f_Z(z + Hu_1) f_Z(z + Hu_2) \\
&\quad \times \; [tr\{H\mathcal{H}_\alpha(z) H u_1 u_1^T\} + tr\{H\mathcal{H}_\alpha(z) H u_2 u_2^T\}] du_1 du_2 \\
&= n(n-1)\tilde{g}(z,z)[f_Z(z)]^2 \mu_2(K) tr\{H^2 \mathcal{H}_\alpha(z)\} + o_p(n^2 tr(H^2));
\end{aligned}$$

and

$$Var\left\{\sum_{i<j}[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{ij}-\mathbf{z})]\right\}$$
$$=O_p(n^2tr(H^2))+O_p(n^3tr(H^3))=o_p(n^4tr(H)^4).$$

Thus

$$\sum_{i<j}[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{ij}-\mathbf{z})]$$
$$=n(n-1)\tilde{g}(z,z)[f_Z(z)]^2\mu_2(K)tr\{H^2\mathcal{H}_\alpha(z)\}+o_p(n^2tr(H^2)).$$

Secondly,

$$\begin{aligned}
&E\left\{\sum_{i<j}[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{ij}-\mathbf{z})](Z_{ij}-\mathbf{z})\right\}\\
&=\frac{n(n-1)}{2}E[\tilde{g}(Z_1,Z_2)K_H(Z_{12}-\mathbf{z})(Z_{12}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{12}-\mathbf{z})(Z_{12}-\mathbf{z})]\\
&=\frac{n(n-1)}{2}\int\int\tilde{g}(z_1,z_2)|H|^{-2}K(H^{-1}(z_1-z))K(H^{-1}(z_2-z))\\
&f_Z(z_1)f_Z(z_2)\left\{((z_1-z)^T,(z_2-z)^T)\mathcal{H}_\alpha(z)\begin{pmatrix}z_1-z\\ z_2-z\end{pmatrix}\right\}\begin{pmatrix}z_1-z\\ z_2-z\end{pmatrix}dz_1dz_2\\
&=\frac{n(n-1)}{2}\int\int K(u_1)K(u_2)g(z+Hu_1,z+Hu_2)f_Z(z+Hu_1)f_Z(z+Hu_2)\\
&[u_1^THH\mathcal{H}_\alpha(z)Hu_1+u_2^THH\mathcal{H}_\alpha(z)Hu_2]\begin{pmatrix}Hu_1\\ Hu_2\end{pmatrix}du_1du_2\\
&=O_p\left[n^2\begin{pmatrix}H^3\mathbf{1}\\ H^3\mathbf{1}\end{pmatrix}\right], \qquad (20)
\end{aligned}$$

where $\mathbf{1}$ denotes the d-dimensional vector with all entries as one. Hence

$$Var\left\{\sum_{i<j}[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{ij}-\mathbf{z})](Z_{ij}-\mathbf{z})\right\}=o_p(n^4H^4).$$

We have

$$\sum_{i<j}[\Delta_{ij}K_H(Z_{ij}-\mathbf{z})(Z_{ij}-\mathbf{z})^T\mathcal{H}_\alpha(z)(Z_{ij}-\mathbf{z})](Z_{ij}-\mathbf{z})=O_p\left[n^2\begin{pmatrix}H^3\mathbf{1}\\ H^3\mathbf{1}\end{pmatrix}\right].$$

Therefore

$$\mathbf{Z}_z^T \mathbf{W}_z \mathbf{Q}_\alpha(z) \\ = \begin{pmatrix} n(n-1)\tilde{g}(z,z)[f_Z(z)]^2 \mu_2(K) tr\{H^2 \mathcal{H}_\alpha(z)\} + o_p(n^2 tr(H^2)) \\ O_p\left[n^2 \begin{pmatrix} H^3\mathbf{1} \\ H^3\mathbf{1} \end{pmatrix}\right] \end{pmatrix}. \quad (21)$$

Appendix 5: Analysis of the matrix $(\mathbf{Z}_z^T \mathbf{W}_z Var(\boldsymbol{\delta}) \mathbf{W}_z \mathbf{Z}_z)$

The matrix $(\mathbf{Z}_z^T \mathbf{W}_z Var(\boldsymbol{\delta}) \mathbf{W}_z \mathbf{Z}_z)$ can be rewritten as

$$\mathbf{Z}_z^T \mathbf{W}_z Var(\boldsymbol{\delta}) \mathbf{W}_z \mathbf{Z}_z = \begin{pmatrix} S_0^* & (S_1^*)^T \\ S_1^* & S_2^* \end{pmatrix},$$

where

$$S_0^* = \sum_{i<j} \sum_{k<l} \Delta_{ij} K_H(Z_{ij} - \mathbf{z}) \Delta_{kl} K_H(Z_{kl} - \mathbf{z}) Cov(\delta_{ij}, \delta_{kl}),$$
$$S_1^* = \sum_{i<j} \sum_{k<l} \Delta_{ij} K_H(Z_{ij} - \mathbf{z}) \Delta_{kl} K_H(Z_{kl} - \mathbf{z}) Cov(\delta_{ij}, \delta_{kl})(Z_{ij} - \mathbf{z}),$$
$$S_2^* = \sum_{i<j} \sum_{k<l} \Delta_{ij} K_H(Z_{ij} - \mathbf{z}) \Delta_{kl} K_H(Z_{kl} - \mathbf{z}) Cov(\delta_{ij}, \delta_{kl})(Z_{ij} - \mathbf{z})(Z_{ij} - \mathbf{z})^T.$$

Notice that $Cov(\delta_{ij}, \delta_{kl}) = 0$ for $i \neq j \neq k \neq l$. So the double summation in $S_k^*, k = 0, 1, 2$ actually involves less than $[n(n-1)/2]^2$ terms, however, there is more than the $n(n-1)/2$ terms for independent δ_{ij}'s. Similar as in Appendix 1, we have the following decomposition

$$\begin{aligned} S_0^* &= \sum_{i<j} \Delta_{ij} K_H^2(Z_{ij} - \mathbf{z}) \tilde{\alpha}(Z_i, Z_j) \\ &\quad + \sum_{i<(j\neq k)} \Delta_{ij} \Delta_{ik} K_H(Z_{ij} - \mathbf{z}) K_H(Z_{ik} - \mathbf{z}) \check{\alpha}(Z_i, Z_j, Z_k) \\ &\quad + \sum_{(i\neq k)<j} \Delta_{ij} \Delta_{kj} K_H(Z_{ij} - \mathbf{z}) K_H(Z_{kj} - \mathbf{z}) \check{\alpha}(Z_j, Z_i, Z_k) \\ &\quad + \sum_{i<j<k} \Delta_{ij} \Delta_{jk} K_H(Z_{ij} - \mathbf{z}) K_H(Z_{jk} - \mathbf{z}) \check{\alpha}(Z_j, Z_i, Z_k) \\ &\quad + \sum_{k<i<j} \Delta_{ij} \Delta_{ki} K_H(Z_{ij} - \mathbf{z}) K_H(Z_{ki} - \mathbf{z}) \check{\alpha}(Z_i, Z_j, Z_k), \end{aligned} \quad (22)$$

where $\check{\alpha}(Z_i, Z_j, Z_k) = Cov[\delta_{ij}, \delta_{ik} | \Delta_{ij} = \Delta_{ik} = 1, Z_i, Z_j, Z_k] = E[\delta_{ij}\delta_{ik} | \Delta_{ij} = \Delta_{ik} = 1, Z_i, Z_j, Z_k] - \tilde{\alpha}(Z_i, Z_j)\tilde{\alpha}(Z_i, Z_k)$. Hence

$$
\begin{aligned}
E[S_0^*|Z_1, ..., Z_n] &= \sum_{i<j} E[K_H^2(Z_{ij} - z)\tilde{g}(Z_i, Z_j)\tilde{\alpha}(Z_i, Z_j)] \\
&+ \sum_{i<(j\neq k)} E[K_H(Z_{ij} - z)K_H(Z_{ik} - z)\breve{g}(Z_i, Z_j, Z_k)\breve{\alpha}(Z_i, Z_j, Z_k)] \\
&+ \sum_{(i\neq k)<j} E[K_H(Z_{ij} - z)K_H(Z_{kj} - z)\breve{g}(Z_j, Z_i, Z_k)\breve{\alpha}(Z_j, Z_i, Z_k)] \\
&+ \sum_{i<j<k} E[K_H(Z_{ij} - z)K_H(Z_{jk} - z)\breve{g}(Z_j, Z_i, Z_k)\breve{\alpha}(Z_j, Z_i, Z_k)] \\
&+ \sum_{k<i<j} E[K_H(Z_{ij} - z)K_H(Z_{ki} - z)\breve{g}(Z_i, Z_j, Z_k)\breve{\alpha}(Z_i, Z_j, Z_k)] \\
&= B_1 + B_2 + B_3 + B_4 + B_5.
\end{aligned} \tag{23}
$$

For the first term,

$$
\begin{aligned}
B_1 &= \frac{n(n-1)}{2} E[K_H^2(Z_{12} - \mathbf{z})\tilde{g}(Z_1, Z_2)\tilde{\alpha}(Z_1, Z_2)] \\
&= \frac{n(n-1)}{2} \int\int \tilde{g}(z_1, z_2)\tilde{\alpha}(z_1, z_2)[|H|^{-2} \\
&\quad \times K(H^{-1}(z_1 - z))K(H^{-1}(z_2 - z))]^2 f_Z(z_1) f_Z(z_2) dz_1 dz_2 \\
&= \frac{n(n-1)}{2}|H|^{-2} \int\int \tilde{g}(z + Hu_1, z + Hu_2)\tilde{\alpha}(z + Hu_1, z + Hu_2) \\
&\quad \times [K(u_1)K(u_2)]^2 f_Z(z + Hu_1) f_Z(z + Hu_2) du_1 du_2 \\
&= \frac{n(n-1)}{2}|H|^{-2}\tilde{g}(z, z)\tilde{\alpha}(z, z)[\mu_0(K^2)]^2[f_Z(z)]^2 + o_p(n^2|H|^{-2}).
\end{aligned}
$$

For the second term,

$$
\begin{aligned}
B_2 &= \sum_{i<(j\neq k)} E[K_H(Z_{ij} - \mathbf{z})K_H(Z_{ik} - \mathbf{z})\breve{g}(Z_i, Z_j, Z_k)\breve{\alpha}(Z_i, Z_j, Z_k)] \\
&= \frac{n(n-1)(n-2)}{4} E[\breve{g}(Z_1, Z_2, Z_3)\breve{\alpha}(Z_1, Z_2, Z_3)K_H(Z_{12} - \mathbf{z})K_H(Z_{13} - \mathbf{z})] \\
&= \frac{n(n-1)(n-2)}{4} \int\int\int \breve{g}(z_1, z_2, z_3)\breve{\alpha}(z_1, z_2, z_3) \\
&\quad \times [|H|^{-2}K(H^{-1}(z_1 - z))K(H^{-1}(z_2 - z))][|H|^{-2}K(H^{-1}(z_1 - z)) \\
&\quad \times K(H^{-1}(z_3 - z))] f_Z(z_1) f_Z(z_2) f_Z(z_3) dz_1 dz_2 dz_3 \\
&= \frac{n(n-1)(n-2)}{4}|H|^{-1}\breve{g}(z, z, z)\breve{\alpha}(z, z, z)\mu_0(K^2)[f_Z(z)]^3 + o_p(n^3|H|^{-1}).
\end{aligned} \tag{24}
$$

Similarly we have

$$B_3 = B_4 = B_5 = \frac{n(n-1)(n-2)}{4}|H|^{-1}\breve{g}(z,z,z)\breve{\alpha}(z,z,z)\mu_0(K^2) \times [f_Z(z)]^3 + o_p(n^3|H|^{-1}).$$

Therefore

$$E[S_0^*|Z_1, ..., Z_n] = n(n-1)(n-2)|H|^{-1}\breve{g}(z,z,z)\breve{\alpha}(z,z,z)\mu_0(K^2) \times [f_Z(z)]^3 + o_p(n^3|H|^{-1}).$$

Since $Var(S_0^*|Z_1, ..., Z_n) = O_p(n^5|H|^{-3})$, we have

$$S_0^* \to n(n-1)(n-2)|H|^{-1}\breve{g}(z,z,z)\breve{\alpha}(z,z,z)\mu_0(K^2)[f_Z(z)]^3.$$

Similarly, we have

$$\begin{aligned} S_1^* &= \sum_{i<j} \Delta_{ij} K_H^2(Z_{ij} - \mathbf{z})\tilde{\alpha}(Z_i, Z_j)(Z_{ij} - \mathbf{z}) \\ &\quad + \sum_{i<(j\neq k)} \Delta_{ij}\Delta_{ik} K_H(Z_{ij} - \mathbf{z})K_H(Z_{ik} - \mathbf{z})\breve{\alpha}(Z_i, Z_j, Z_k)(Z_{ij} - \mathbf{z}) \\ &\quad + \sum_{(i\neq k)<j} \Delta_{ij}\Delta_{kj} K_H(Z_{ij} - \mathbf{z})K_H(Z_{kj} - \mathbf{z})\breve{\alpha}(Z_j, Z_i, Z_k)(Z_{ij} - \mathbf{z}) \qquad (25) \\ &\quad + \sum_{i<j<k} \Delta_{ij}\Delta_{jk} K_H(Z_{ij} - \mathbf{z})K_H(Z_{jk} - \mathbf{z})\breve{\alpha}(Z_j, Z_i, Z_k)(Z_{ij} - \mathbf{z}) \\ &\quad + \sum_{k<i<j} \Delta_{ij}\Delta_{ki} K_H(Z_{ij} - \mathbf{z})K_H(Z_{ki} - \mathbf{z})\breve{\alpha}(Z_i, Z_j, Z_k)(Z_{ij} - \mathbf{z}) \\ &= O_p[n^2|H|^{-2}\begin{pmatrix} H\mathbf{1} \\ H\mathbf{1} \end{pmatrix}] + O_p[n^3|H|^{-1}\begin{pmatrix} H\mathbf{1} \\ H\mathbf{1} \end{pmatrix}], \end{aligned}$$

$$\begin{aligned} S_2^* &= \sum_{i<j} \Delta_{ij} K_H^2(Z_{ij} - \mathbf{z})\tilde{\alpha}(Z_i, Z_j)(Z_{ij} - \mathbf{z})(Z_{ij} - \mathbf{z})^T \\ &\quad + \sum_{i<(j\neq k)} \Delta_{ij}\Delta_{ik} K_H(Z_{ij} - \mathbf{z})K_H(Z_{ik} - \mathbf{z})\breve{\alpha}(Z_i, Z_j, Z_k)(Z_{ij} - \mathbf{z})(Z_{ik} - \mathbf{z})^T \\ &\quad + \sum_{(i\neq k)<j} \Delta_{ij}\Delta_{kj} K_H(Z_{ij} - \mathbf{z})K_H(Z_{kj} - \mathbf{z})\breve{\alpha}(Z_j, Z_i, Z_k)(Z_{ij} - \mathbf{z})Z_{kj} - \mathbf{z})^T \\ &\quad + \sum_{i<j<k} \Delta_{ij}\Delta_{jk} K_H(Z_{ij} - \mathbf{z})K_H(Z_{jk} - \mathbf{z})\breve{\alpha}(Z_j, Z_i, Z_k)(Z_{ij} - \mathbf{z})(Z_{jk} - \mathbf{z})^T \\ &\quad + \sum_{k<i<j} \Delta_{ij}\Delta_{ki} K_H(Z_{ij} - \mathbf{z})K_H(Z_{ki} - \mathbf{z})\breve{\alpha}(Z_i, Z_j, Z_k)(Z_{ij} - \mathbf{z})(Z_{ki} - \mathbf{z})^T \\ &= O_p\left[n^2|H|^{-2}\begin{pmatrix} H^2 & 0 \\ 0 & H^2 \end{pmatrix}\right] + O_p\left[n^3|H|^{-1}\begin{pmatrix} H^2 & H^2 \\ H^2 & H^2 \end{pmatrix}\right]. \qquad (26) \end{aligned}$$

Therefore

$$\mathbf{Z}_z^T \mathbf{W}_z Var(\boldsymbol{\delta}) \mathbf{W}_z \mathbf{Z}_z \to n(n-1)(n-2)|H|^{-1} \breve{g}(z,z,z) \breve{\alpha}(z,z,z) \mu_0(K^2)[f_Z(z)]^3 \times \begin{pmatrix} 1 & O_p[H\mathbf{1}]^T & O_p[H\mathbf{1}]^T \\ O_p[H\mathbf{1}] & O_p[H^2] & O_p[H^2] \\ O_p[H\mathbf{1}] & O_p[H^2] & O_p[H^2] \end{pmatrix}.$$

Appendix 6: Proof of Theorem 2

Here we modify the proof of Theorem 1 similar to the proof of Theorem 2.2 in Ruppert and Wand (1994). The key point for boundary treatment is to write the estimator (10) in terms of the equivalent kernel $K^*(u; z) = e_1^T (\mathbf{Z}_z^T \mathbf{W}_z \mathbf{Z}_z)^{-1} (1\ (u-z)^T)^T K_H(u-z)$. Then $\hat{\alpha}(z) = \sum_{i<j} K^*(Z_{ij}; z) \Delta_{ij} \delta_{ij}$,

$$\sum_{i<j} K^*(Z_{ij}; z)\Delta_{ij} = 1, \qquad \sum_{i<j} K^*(Z_{ij}; z)\Delta_{ij}(Z_{ij} - z) = 0. \tag{27}$$

The moment condition (27) ensures the asymptotic conditional bias and variance are given by (11) and (12) at both the interior and the boundary. Then we only need to check the derivations in Appendices C to E to get different approximations for matrices $\mathbf{Z}_z^T \mathbf{W}_z \mathbf{Z}_z$ and $\mathbf{Z}_z^T \mathbf{W}_z Var(\boldsymbol{\delta}) \mathbf{W}_z \mathbf{Z}_z$. At boundary points, the integral of kernel function $K(u)$ over the whole range would be replaced by its integral over only the region $\mathcal{D}_{z,H}$. Thus the quantities in Theorem 1 are modified accordingly.

Quick examination of Eqs. (15)–(18) shows that, at boundary point, Eq. (13) should change into

$$\mathbf{Z}_z^T \mathbf{W}_z \mathbf{Z}_z = \frac{n(n-1)}{2} \tilde{g}(z,z)[f_Z(z)]^2 \tilde{H} N_z \tilde{H}[1 + o_p(1)], \tag{28}$$

with $\tilde{H}$ defined in (14).

Examination of Eqs. (19) and (20) shows that, at boundary point, $\mathbf{Z}_z^T \mathbf{W}_z \mathbf{Q}_\alpha(z)$ in Eq. (21) converges to

$$\frac{n(n-1)}{2} \tilde{g}(z,z)[f_Z(z)]^2 \times \begin{pmatrix} \int_{\mathcal{D}_{z,H}} \int_{\mathcal{D}_{z,H}} K(u_1)K(u_2)[u_1^T H \mathcal{H}_\alpha(z) H u_1 + u_2^T H \mathcal{H}_\alpha(z) H u_2] du_1 du_2 \\ \int_{\mathcal{D}_{z,H}} \int_{\mathcal{D}_{z,H}} H u_1 K(u_1)K(u_2)[u_1^T H \mathcal{H}_\alpha(z) H u_1 + u_2^T H \mathcal{H}_\alpha(z) H u_2] du_1 du_2 \\ \int_{\mathcal{D}_{z,H}} \int_{\mathcal{D}_{z,H}} H u_2 K(u_1)K(u_2)[u_1^T H \mathcal{H}_\alpha(z) H u_1 + u_2^T H \mathcal{H}_\alpha(z) H u_2] du_1 du_2 \end{pmatrix}$$

$$= \frac{n(n-1)}{2} \tilde{g}(z,z)[f_Z(z)]^2 \tilde{H} \begin{pmatrix} 2\mu_{z,0}^*(K)\mu_{z,2}^H(K) \\ \mu_{z,2}^H(K)\mu_{z,1}^*(K) + \mu_{z,0}^*(K)\mu_{z,3}^H(K) \\ \mu_{z,2}^H(K)\mu_{z,1}^*(K) + \mu_{z,0}^*(K)\mu_{z,3}^H(K) \end{pmatrix}.$$

Plug this and (28) into (11), we get (6).

Examination of Eqs. (22)–(26) shows that, at boundary point, $\mathbf{Z}_z^T \mathbf{W}_z Var(\boldsymbol{\delta}) \mathbf{W}_z \mathbf{Z}_z$ converges to

$$n(n-1)(n-2)|H|^{-1} \check{g}(z,z,z) \check{\alpha}(z,z,z) [f_Z(z)]^3 \tilde{H} T_z \tilde{H} [1+o_p(1)].$$

Plug this and (28) into (12) we get (7)

$$Var[\hat{\alpha}(z)|Z_1, ..., Z_n] = \frac{4\check{g}(z,z,z)\check{\alpha}(z,z,z) e_1^T N_z^{-1} T_z N_z^{-1} e_1}{n|H|[\tilde{g}(z,z)]^2 f_Z(z)} [1+o_p(1)].$$

References

Acar EF, Craiu RV, Yao F (2011) Dependence calibration in conditional copulas: a nonparametric approach. Biometrics 67:445–453

Clayton DG (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65:141–51

Chaieb L, Rivest L-P, Abdous B (2006) Estimating survival under a dependent truncation. Biometrika 93:655–669

Day R, Bryant J, Lefkopolou M (1997) Adaptation of bivariate frailty models for prediction, with applications to biological markers as prognostic indicators. Biometrika 84:45–56

Ding AA (2010) Identifiability conditions for covariate effects model on survival times under informative censoring. Stat Probab Lett 80:911–915

Ding AA (2012) Copula identifiability conditions for dependent truncated data model. Lifetime Data Anal 18(4):397–407

Ding AA, Shi G, Wang W, Hsieh JJ (2009) Marginal regression analysis for semi-competing risks data under dependent censoring. Scand J Stat 36:481–500

Emura T, Wang W (2010) Semi-parametric inference for copula models for truncated data. Stat Sinica 21:349–367

Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Chapman and Hall, Boca Raton

Fine JP, Jiang H (2000) On sssociation in a copula with time transformations. Biometrika 87:559–571

Fine JP, Jiang H, Chappell R (2001) On semi-competing risks data. Biometrika 88:907–919

Gijbels I, Veraverbeke N, Omelka M (2011) Conditional copulas, association measures and their applications. Comput Stat Data Anal 55:1919–1932

Ghosh D (2006) Semiparametric inferences for association with semi-competing risks data. Stat Med 25:2059–2070

Hsieh J, Wang Ding AA (2008) Regression analysis based on semicompeting risks data. J R Stat Soc Ser B 70:3–20

Huang X, Zhang N (2008) Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. Biometrics 64:1090–1099

Jiang H, Fine JP, Chappell R (2005) Semiparametric analysis of survival data with left truncation and dependent right censoring. Biometrics 61:567–575

Klein JP, Moeschberger ML (2003) Survival analysis: techniques for censored and truncated data, 2nd edn. Springer, New York

Lakhal L, Rivest L-P, Abdous B (2008) Estimating survival and sssociation in a semicompeting risks model. Biometrics 64:180–188

Lin DY, Robins JM, Wei LJ (1996) Comparing two failure time distributions in presence of dependent censoring. Biometrika 83:381–393

Lin DY, Ying Z (2003) Semiparametric regression analysis of longitudinal data with informative drop-outs. Biostatistics 4:385–398

Martin EC, Betensky RA (2005) Testing quasi-independence of failure and truncation via Conditional Kendalls Tau. J Am Stat Assoc 100:484–492

Nelsen RB (2006) An introduction to copulas. Springer, New York

Oakes D (1989) Bivariate survival models induced by frailties. J Am Stat Assoc 84:487–493

Opsomer J-D, Ruppert D (1997) Fitting a bivariate additive model by local polynomial regression. Ann Stat 25:186–211
Peng L, Fine JP (2006) Rank estimation of accelerated lifetime models with dependent censoring. J Am Stat Assoc 101:1085–1093
Peng L, Fine JP (2007) Regression modeling of semicompeting risks data. Biometrics 63:96–108
Ruppert D, Wand MP (1994) Multivariate locally weighted least squares regression. Ann Stat 22:1346–1370
Shih JH, Louis TA (1995) Inference on the association parameter in copula models for bivariate survival data. Biometrics 51:1384–99
Tsai WY (1990) Testing the assumption of independence of truncation time and failure time. Biometrika 77:169–177
Wang W (2003) Estimation of the association parameter for copula models under dependent censoring. J R Stat Soc Ser B 65:257–273
Wang W, Wells MT (2000) Model selection and semiparametric inference for bivariate failure time data. J Am Stat Assoc 95:62–72